

A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics

Southam, Andrew D; Weber, Ralf J M; Engel, Jasper; Jones, Martin R; Viant, Mark R

DOI:

[10.1038/nprot.2016.156](https://doi.org/10.1038/nprot.2016.156)

License:

None: All rights reserved

Document Version

Peer reviewed version

Citation for published version (Harvard):

Southam, AD, Weber, RJM, Engel, J, Jones, MR & Viant, MR 2017, 'A complete workflow for high-resolution spectral-stitching nanoelectrospray direct-infusion mass-spectrometry-based metabolomics and lipidomics', *Nature protocols*, vol. 12, no. 2, pp. 310-328. <https://doi.org/10.1038/nprot.2016.156>

[Link to publication on Research at Birmingham portal](#)

Publisher Rights Statement:

Final Version of Record published as above.

Checked 24/2/2017

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

A complete workflow for High-Resolution Spectral-Stitching Nanoelectrospray Direct Infusion Mass Spectrometry-based Metabolomics and Lipidomics

Andrew D. Southam^{1#}, Ralf J. M. Weber^{1#}, Jasper Engel², Martin R. Jones¹ and Mark R. Viant^{1,2*}

¹School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

²NERC Biomolecular Analysis Facility – Metabolomics Node (NBAF-B), School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[#]these authors contributed equally to this study

* Telephone +44 (0)121 4142219, Fax +44 (0)121 4145925 Website:

<http://www.birmingham.ac.uk/schools/biosciences/staff/profile.aspx?ReferenceId=9910&Name=professor-mark-viant>

ABSTRACT

Metabolomics and lipidomics measure and discover metabolic and lipid profiles in biological samples. These technologies thereby enable the metabolism of specific biological phenotypes to be understood. Accurate biological interpretations require high analytical reproducibility and sensitivity, and standardised and transparent data processing. Here we describe a complete workflow for nanoelectrospray ionisation (nESI) direct infusion mass spectrometry (DIMS) metabolomics and lipidomics. After metabolite and lipid extraction from tissues and biofluids, samples are directly infused into a high-resolution mass spectrometer (e.g. Orbitrap) using a chip-based nESI sample delivery system. nESI functions to minimise ionisation suppression or enhancement effects compared to standard electrospray ionisation. Our analytical technique—named spectral-stitching—measures data as several overlapping mass-to-charge (m/z) windows that are subsequently 'stitched' together creating a complete mass spectrum. This considerably increases dynamic range and detection sensitivity—ca. 5-fold increase in peak detection—compared to the collection of DIMS data as a single wide m/z window. Data processing, statistical analysis and metabolite annotation are executed as a workflow within the user-friendly, transparent and freely available Galaxy platform (galaxyproject.org). Generated data have high mass accuracy that enables molecular formulae peak annotations. The workflow is compatible with any sample extraction method; in this protocol the examples are extracted using a biphasic method with methanol, chloroform and water as the solvents. The complete workflow is reproducible, rapid and automated, which enables cost-effective analysis of

>10,000 samples per year making it ideal for high-throughput metabolomics and lipidomics screening, e.g. for clinical phenotyping, drug screening and toxicity testing.

INTRODUCTION

Non-targeted metabolomics and lipidomics techniques comprise the measurement and analysis of the steady-state levels of the multiple metabolites or lipids (termed the metabolome or lipidome) within biological samples¹⁻⁴. Comparison of the metabolome or lipidome across different phenotypes represents a powerful and unbiased approach to discover molecular perturbations that can in turn be used to generate biological hypotheses for more detailed investigation. This approach is applicable to diverse sample types, including biofluids^{5,6}, mammalian cells⁷ and spent cell media⁸, tissues⁹ and whole organisms¹⁰, and to a wide range of biological questions such as the investigation of disease⁹, drug action⁷ and toxicology¹¹. Metabolites and lipids have many diverse functions in biological systems. They are typically the end products of complex cellular regulation networks¹ (at the genetic, epigenetic, transcriptional, translational and post-translational levels) and they can also influence and alter this regulation via feedback loops¹², protein modifications¹³ and epigenetic changes¹⁴. They are both building blocks of complex cellular macromolecules and also sources and intermediates in energy metabolism. Thus, metabolomics and lipidomics analyses can give insightful knowledge of the functional molecular status of biological systems (e.g. energetic status and the balance of anabolic and catabolic processes) and can complement other 'omic technologies as well as traditional molecular biological investigations.

To accurately and reliably interpret data derived from metabolomics and lipidomics studies, the entire workflow including the experimental design, sample collection and extraction, as well as data acquisition, processing, metabolite identification and statistical analysis, should be robust and reproducible. High detection sensitivity of the chosen analytical approach is also desirable. The most common analytical techniques used in metabolomics include nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS)¹⁵. While NMR is highly quantitative and reproducible, MS methods are increasingly favoured for metabolomics and lipidomics due to their higher sensitivity^{15,16}. MS approaches often include chromatographic separation that serves to resolve the complex mixture of metabolites or lipids prior to ion detection, thereby helping to distinguish between isobaric compounds and to minimise ionisation suppression (or enhancement) effects. It is because of these attributes that gas chromatography (GC) MS¹⁷ and liquid chromatography (LC) MS¹⁸, and to a lesser extent capillary electrophoresis (CE) MS¹⁹ and ion chromatography (IC) MS²⁰, are popular and widely employed for metabolomics and lipidomics investigations. Direct infusion mass spectrometry (DIMS) is an alternative approach involving the direct introduction of biological extracts into MS systems without any prior chromatographic separation²¹. Nominal mass, flow injection electrospray ionisation (ESI) DIMS approaches have previously been developed and successfully applied to high throughput metabolomics²². However, significantly higher quality DIMS data can be acquired by employing automated nanoelectrospray ionisation (nESI) sample delivery platforms in combination with ultra-high mass resolution and accuracy MS detectors (e.g. Fourier transform ion cyclotron resonance (FT-

ICR) MS, Orbitrap MS)^{10,23,24}. These platforms increase the number of detectable peaks by minimising ionisation suppression and enhancement effects (relative to ESI)²⁵. They also maximise the discrimination of peaks with similar accurate masses, thereby facilitating more accurate molecular formula(e) peak annotations of the data²⁶.

The protocol described here is a complete workflow for conducting high-resolution nESI DIMS metabolomics and lipidomics investigations (Figure 1), guided by the Metabolomics Standards Initiative framework²⁷. The optimised nESI DIMS method, termed 'spectral-stitching nESI DIMS', collects MS data as a series of overlapping mass-to-charge (m/z) windows that are subsequently stitched together into a complete spectrum (Figure 2, Table 1). This approach maximises both the quantity and mass accuracy of detectable peaks^{23,24}. The automated data processing and analysis steps are collated into an open source workflow within Galaxy²⁸. This offers a standardised, transparent and user-friendly approach without the requirement for bioinformatics expertise and knowledge of multiple programming languages and/or environments. Galaxy is an open source workflow platforms that is used for Next Generation Sequencing (NGS) data analysis. It has many standard processing tools (accessible from its web-based user interface) that improve both the speed and reproducibility of data processing and analysis. Our recent development of Galaxy-M spectral-stitching nESI DIMS workflow considerably increases the accessibility of the spectral-stitching nESI DIMS metabolomics approach²⁸.

Advantages and limitations of the spectral-stitching nESI DIMS workflow

A major advantage of the spectral-stitching nESI DIMS workflow is the generation of high quality metabolomics or lipidomics data, with high spectral resolution, mass accuracy and dynamic range, in a short timeframe and cost-effective manner. The high resolution and mass accuracy allows for the accurate annotation of peaks with molecular formula(e) and for compounds with similar masses to be resolved^{23,26}, which is particularly advantageous when analysing complex mixtures of chemicals. Spectral-stitching nESI DIMS increases detection sensitivity 5-fold compared to standard wide-scan DIMS (for a constant number of ions entering the detector each scan)²³ and is highly reproducible: the median relative standard deviation (RSD) of peak intensities in 80 repeated injections of the same biological sample was 8.2%²⁹. The intensity measurements are robust and correlate well with peak intensities measured using quantitative NMR methods for selected metabolites^{11,30}, demonstrating the high relative-quantification capability of the method (see *ANTICIPATED RESULTS*, below). Additionally, DIMS has been suggested to have comparable metabolomic classification and prediction capabilities as LC-MS³¹. DIMS is more applicable for lipidomics than NMR due to the highly-congested lipid region in ¹H NMR spectra, and the higher sensitivity of DIMS compared to NMR is a clear advantage for both metabolomics and lipidomics. Compared to chromatography-based MS, DIMS approaches — including the spectral-stitching method — have considerably shorter acquisition times (enabling higher throughput); no chromatographic drift (as observed for LC-MS as the column

ages) enabling more robust alignment of multiple spectra; no carry-over between biological samples; and lower consumables costs.

The spectral-stitching nESI DIMS workflow is a global profiling method and therefore cannot detect all metabolites or lipids in a single experiment. For the analysis of specific metabolites or lipids, extraction procedures designed to maximise these compound classes should be employed prior to DIMS data collection. A limitation of DIMS is ion suppression or ion enhancement, which is caused by multiple different compounds entering the ionisation source at the same time³². However, the low flow rate associated with nESI, as used here, minimises these artefacts^{25,33}. Sample lipid concentration can vary between sample types. To prevent high abundant lipids dominating the DIMS analysis, a dilution study is required to find the optimal sample dilution (see *Resuspension of biological sample extracts in DIMS solvents* in the protocol below)—a sample lipid concentration of < 10 pmol/μL is recommended³⁴. The limitations of DIMS compared to chromatography-based MS or NMR techniques include the inability to resolve isobaric compounds and the inability to provide any structural information on compounds unless further MS/MS analyses are conducted. However it is typically very difficult to isolate just the ion of interest for DIMS/MS given the complexity of the mass spectra¹¹. A further limitation of DIMS is the relative-quantitative nature of the measurement, i.e. as for LC-MS metabolomics it measures relative fold-changes between biological samples. In comparison, NMR can be regarded as fully quantitative and LC-MS/MS using internal standards is also quantitative, albeit just for selected peaks of interest. If desired, these approaches can be used to fully quantify metabolites that are discovered to be important by the spectral-stitching nESI DIMS method workflow. Therefore DIMS, chromatography-based MS and NMR each possess different advantages and limitations and act as complementary approaches for metabolomics and lipidomics dependent upon the analysts requirements.

Applications of the spectral-stitching nESI DIMS workflow

The spectral-stitching nESI DIMS workflow is ideal for large scale discovery studies²⁹. It has successfully been applied to a diverse range of biological sample types, including tissues¹¹, mammalian cells⁷, whole organism homogenates¹⁰ and haemolymph³⁵, and is also applicable to urine³⁶, serum³¹ and plasma analysis³⁷. Furthermore it has been used to investigate a broad range of biological questions including studies into xenobiotic toxicity in whole organisms¹⁰ and in specific organs^{11,38}; identification of metabolic changes associated with disease³⁹; elucidation of the action of therapeutic drugs on the lipidome of human leukaemia cells⁷; and metabolic footprinting analysis to understand the interactions of algae with their environment⁴⁰. The high resolution of the MS detector means that this workflow is also compatible with ¹³C-labelling ‘pulse-chase’ or ‘flux’ approaches, wherein the metabolism and fate of a ¹³C-labeled substrate such as ¹³C-glucose or ¹³C-glutamine is measured as a function of time, e.g., measurement of ¹³C-glucose incorporation into fatty acids and phospholipids in response to drug treatments⁷. A further example of the application of this DIMS workflow is for the classification of compounds as being of endogenous, exogenous or metabolised-

exogenous origin, following exposure of an organism to an undefined chemical mixture, by analysing both the organism's tissue and the exposome⁴¹. Given that metabolomics has now matured to a widely used, relatively stable analytical and computational approach, with a rapidly growing community supported by a wide range of training opportunities for researchers, the field is now poised to tackle very large scale challenges in the biomedical and regulatory sciences. As one example, regulatory toxicology is now opening the door to new testing strategies for determining the impacts of chemicals on human health. Given that there are approaching 100,000 chemicals used in industry and consumer products, the need for very high throughput screening approaches, incorporating 'omics data generation, are paramount. The protocol reported here is an important step towards translating metabolomics into regulatory toxicology^{42,43}.

Experimental design

A well-designed experiment is essential to ensure a meaningful and robust outcome of a DIMS study. Biological sample collection, preparation and nESI DIMS data generation must be randomised across sample classes (typically different biological phenotypes) to prevent user induced bias. The amount of biological replication required to provide adequate statistical power to the study will be influenced by the type of experiment being conducted. Based upon a decade of expertise, our recommended levels of biological replication vary from n=6 (for well controlled laboratory based studies with relatively little inter-sample biological variance, such as nESI DIMS of mammalian cell culture extracts) to n=10 (controlled laboratory studies but with greater biological variance, such as model organism studies) to larger n for cases where biological variance is not controlled (clinical and environmental biological samples). With pilot data this estimation can be improved by applying sample size and power analyses (MetaboAnalyst 3.0, <http://www.metaboanalyst.ca>). Technical replication should also be included in the nESI DIMS workflow as a method for distinguishing peaks of biological origin from background noise, increasing the accuracy of the intensity measurements and allowing for an estimation of the reproducibility of peak intensity measurements⁴⁴ (see the *Data processing* section, below). To distinguish peaks of biological origin from genuine but non-biological peaks, such as may arise from the sample preparation method (i.e. from solvents and/or contaminants), an 'extract blank' (sample preparation procedure conducted in the absence of a biological sample) should also be prepared and analysed. Quality control (QC) samples are a critical component of both the nESI DIMS experimental design and data quality control (see *Data Quality Control* section, below). Internal QC samples should be derived by pooling a small volume of ideally all biological samples within one metabolomics study, and are then analysed at regular intervals throughout the analytical run, e.g. every fifth or sixth DIMS analysis is of a QC sample. Internal QC samples are used to monitor and correct for slight analytical variance between samples acquired through one or more analytical batches within that one study (see the *signal correction algorithm*, *Data processing* section). External QC samples (i.e. using a surrogate sample) can be used to monitor and correct for analytical variance between different metabolomics studies.

Biological sample collection, storage and preparation

The time between biological sample collection and freezing should be as short as possible to rapidly halt enzymatic activity and cease changes to the metabolome or lipidome. For mammalian cell culture, a quenching approach is used where cells are sampled directly into a quenching solution (such as -40 °C aqueous methanol)^{7,45}. For the collection of small model organisms (e.g. *Daphnia*, *Drosophila*)⁴⁶, tissue from larger organisms⁴⁷ and biofluids (e.g. urine⁴⁸, serum⁴⁹ or plasma⁴⁹) we refer the reader to the appropriate specialist literature. Biological samples must remain frozen at -80 °C until extraction of the metabolites and/or lipids.

Preparation of solid biological samples typically involves their physical disruption (by a homogenisation probe, e.g. Polytron; or a bead-based homogeniser, e.g. Precellys@24⁵⁰) in the presence of solvents to simultaneously extract the polar metabolites and/or lipids into a liquid phase, while denaturing the metabolic enzymes to halt any further changes to the metabolome or lipidome^{50,51}. For serum, plasma⁴⁹ and mammalian cell suspensions⁷, vortex mixing in the presence of the extraction solvent is sufficient to denature enzymes and to extract the metabolites and/or lipids. Urine can be more simply prepared by centrifugation and dilution with water due to its very low protein (and therefore very low enzyme) content⁴⁸. Extraction techniques should be reproducible, i.e. introduce minimal technical variation to the study, and should achieve as high a yield of metabolites and/or lipids as possible. Several solvent systems can be used^{50,51}. Fine adjustments to the extraction conditions, e.g. pH and solvent polarity, can be used to extract specific classes of compounds according to their solubility characteristics⁵². Bi-phasic extraction methods employ immiscible solvents to simultaneously extract polar metabolites and lipids into separate phases^{50,51,53,54}. This reduces biological sample complexity, which is advantageous for (non-chromatography based) nESI DIMS. Our optimised bi-phasic extraction—based on a procedure developed by Bligh and Dyer⁵³—uses a methanol:chloroform:water ratio of 2:2:1.8 to maximise metabolite yield, metabolic sample stability and reproducibility⁵⁰. Alternatively, the use of methyl tert-butyl ether (MTBE), methanol and water is an increasingly popular bi-phasic approach for lipidomics that improves the extraction of some lipid classes (e.g. ceramides) compared to the Bligh and Dyer method⁵⁴.

Following extraction, extracts should be dried to maintain stability and, subsequently, to enable resuspension in appropriate solvents for DIMS; note that lipids must be dried and stored under nitrogen to prevent oxidative lipid damage, while polar metabolites in an aqueous methanol solution are more typically dried using a vacuum centrifugation system, such as the SpeedVac concentrator.

Direct infusion mass spectrometry

Sample ionisation and analytical detection methods can strongly influence the quality of DIMS metabolomics and lipidomics data. Electrospray ionisation (ESI), a soft-ionisation technique, is

particularly appropriate for DIMS since it can predominantly ionise the metabolites and/or lipids as intact compounds¹⁶ aiding compound molecular formula(e) annotations. A very low flow rate through the electrospray ionisation source is essential for minimising both ionisation suppression and enhancement effects^{25,33}; our nanoflow ESI methods typically use a flow rate of 200-300 nL/min compared to the several μ L/min employed for standard ESI sources. nESI also benefits from lower sample volume requirements and considerably higher sensitivity⁵⁵. To aid nESI compound ionisation, a modifying agent is typically added to the solubilised biological sample (e.g. formic acid or ammonium acetate), which will also influence the types of ion forms that are created (see the *Peak annotation* section, below). High lipid concentration in samples can lead to aggregate formation (e.g. dimers) during DIMS analysis³⁴. This can be prevented by keeping lipid concentrations < 10 pmol/ μ L when using 2:1 methanol:chloroform³⁴. For high throughput metabolomics and lipidomics, an automated chip-based nESI source (e.g. Triversa Nanomate, Advion Biosciences) enables rapid, reproducible and automated acquisition of multiple biological samples without sample-to-sample cross contamination^{25,55}.

Ultra-high resolution, high mass accuracy detectors (e.g. Orbitrap series of mass spectrometers, FT-ICR, both with a resolution capable of exceeding 100k at m/z 200) are ideal for DIMS as they resolve peaks with similar accurate masses and allow for the annotation of peaks with molecular formula(e). This increases the coverage of the metabolome or lipidome and enhances annotation confidence compared to lower resolution, lower mass accuracy instruments. For Thermo Scientific Orbitrap series and FT-ICR mass spectrometers, the number of ions entering the detector is controlled by the Automatic Gain Control (AGC). The number of ions within the detector is approximately proportional to detection sensitivity, however if the number of ions is too high then the mass accuracy decreases (due to space-charge effects⁵⁶), ultimately making unique molecular formula(e) annotation impossible. This is particularly problematic when collecting DIMS data over a single large mass-to-charge range (e.g. 100-1000 m/z) for which a higher AGC setting is required to capture and then detect the lower abundance compounds. A proven solution to this is provided by the spectral-stitching nESI method where data is collected as a series of overlapping m/z windows — by selected ion monitoring (SIM) scan mode for the Thermo Scientific FT-ICR and Orbitrap series — and then ‘stitching’ these windows together to create a complete mass spectrum. The AGC target value does not need to be set as high for narrower m/z windows (with fewer detectable features) compared to a full scan spectrum to achieve similar peak detection sensitivity. Overall, this approach increases the dynamic range and detection sensitivity, while retaining high mass accuracy by minimising space-charge effects that would arise from high AGC settings (Figure 2, Table 1)^{23,24}.

The spectral-stitching method was optimised for the FT-ICR, Q ExactiveTM and Orbitrap EliteTM DIMS by determining the highest AGC setting (i.e. highest sensitivity) that retains high mass accuracy; by establishing the m/z window width for maximal peak detection; and by calculating the minimum number of scans for reliable and reproducible peak detection. Detector sensitivity at each end of the individual narrow m/z windows was found to be lower than across the majority of each m/z range and thus was characterised to establish the usable high sensitivity region of each m/z window that was

retained (and hence determine the overlap of the m/z windows that was required in the method to prevent signal loss). Prior to the application of this method, optimisation of the biological sample concentration is highly advisable in order to prevent poor nESI stability, poor RSDs of the peak intensities from samples that are too concentrated, and poor detection sensitivity for those samples that are not sufficiently concentrated.

Data processing

Prior to data processing, DIMS spectra should be viewed using the vendor's software to identify any analyses for which the electrospray failed during acquisition, and such analyses are removed. DIMS data on both FT-ICR and Orbitrap platforms are recorded as transient (time domain) files as triplicate technical analyses^{23,24} (Figure 1); for some mass spectrometers time domain data is only available to the analyst as a single pre-processed vendor-encoded file (.RAW file in the case of Thermo Scientific), while for other spectrometers the data is available in its rawest form as multiple transient data files. For this latter case (for Bruker and Thermo Scientific FT-ICR MS), data processing includes averaging of the transient data, Hanning apodisation and zero-filling, followed by Fourier transformation to convert time into the frequency domain using custom written Matlab code^{23,24,57}. From this point, data processing is consistent across all DIMS approaches described in this protocol (Figure 1). To remove obvious noise features from the data, any peaks with a signal-to-noise ratio (SNR) threshold of less than typically 3:1 are removed. Data are then mass calibrated to convert from the frequency domain into m/z values, utilising a calibration equation [for the FT-ICR, Thermo Scientific use $m/z = (A/f) + (B/f^2)$ ^{23,56}; for the Orbitrap instrument, $m/z = (B/f^2) + (C/f^4)$ ^{58,59}; in both cases f is the frequency and A, B and C are calibration parameters from the instrument]. External calibration utilises the calibration parameters derived from the periodic calibration of the mass spectrometer using a defined sample of known standards. Post-acquisition internal calibration is used to increase the mass accuracy in each window by using the accurate masses of metabolites known to be present in the mass spectra of the biological samples. Finally the multiple narrow m/z windows are stitched together into one continuous m/z spectrum using custom written Matlab code.

To identify genuine peaks and remove noise features within the mass spectrum of each biological sample, only peaks present in at least two-out-of-three of the technical replicate analyses of the sample are retained (termed 'replicate filtering'). Contaminate peaks arising from the extraction method are flagged by comparing the biological sample spectra to the extract blank, and are then removed if their intensity in the sample is <3-times that of the extract blank (termed 'blank filtering'). The m/z data from multiple biological samples are combined into a single data matrix with samples as rows and m/z features as columns. Peaks are retained if they are present in a given percentage of samples defined by the user (termed 'sample filtering'), e.g. 100% for the calculation of normalisation quotients (see below); >80% as a default setting to achieve robust sample filtering⁵⁷; and lower percentage filtering where multiple classes exist in the dataset and peaks within a single class need to be retained. Note that the lower the sample filter target percentage the greater the number of missing values will be created in the dataset. At this point, the number of peaks within each biological sample

is assessed, and those with abnormally high numbers of missing values are deemed technically poor and removed using 'missing-value filtering' (see *Data Quality Control* section, below).

Prior to statistical analysis several data pre-processing steps are required. As a first step data matrices are normalised by the probabilistic quotient normalisation (PQN) approach to remove the variation caused by unequal amounts of detectable metabolites in each of the biological samples⁶⁰. For studies in which the analytical measurements are made across several batches, the intensity measurements of QC peaks in different batches are normalised to correct the technical variation arising from inter-batch measurements using a *signal correction* method⁶¹. This approach can also be used to correct signal-intensity technical variation that arises within a single batch. The *signal correction* method requires QC samples to be acquired at regular intervals throughout the analytical run. Next, the relative standard deviations (RSDs) of the intensities of each spectral peak are assessed within the QC samples⁴⁴, and any peaks with unacceptably high RSDs are deemed unreliable measurements and removed using the 'peak quality filter' (see *Data Quality Control* section, below). Missing values are now imputed into the normalised DIMS data matrices. Missing values generally occur because: (i) the metabolite is not present or (ii) the instrument fails to detect a metabolite that was present. As missing values can be problematic for statistical analyses, they are typically imputed using the K-nearest neighbour method (KNN), which has been demonstrated to be the most suitable for DIMS data⁶². However, in cases where the user deems missing values to arise from genuinely missing metabolites they may choose to omit this step. Prior to multivariate statistics, a generalised log transformation is optimised (for each specific dataset using the QC sample data) and applied to stabilise the technical variance and reduce the dominance of the highly intense and variable peaks in the multivariate statistics^{63,64}.

Data Quality Control

In summary, seven filtering steps are applied in the pipeline to ensure high data quality:

- 1) if the nESI spray current fails during data collection on the mass spectrometer, the affected analyses are manually discarded;
- 2) for peaks to be considered as actual detected features, they must be observed in at least two DIMS measurements of the biological sample (termed 'replicate filtering', see above and Step 20);
- 3) only peaks occurring at more than 3-times higher intensity in the biological samples relative to the blank samples are retained in the dataset (termed 'blank filtering', see above and Step 23);
- 4) all peaks retained in the dataset occur in the majority of biological samples measured in the study (termed 'sample filtering', see above and Step 24);

- 5) all biological samples have a relatively consistent number of peaks (termed 'missing-value filtering', see above and Step 25);
- 6) any batch effects or temporal drifts in peak intensity, assessed on a peak-by-peak basis using the QC samples, are corrected for (termed 'signal correction', see above and Step 28);
- 7) the reproducibility of peak intensities are calculated within the QC samples across the dataset on a peak-by-peak basis. Non-reproducibly measured peaks are excluded (termed 'peak quality filter', see Step 28)

Univariate and multivariate statistical analysis

Non-targeted nESI DIMS metabolomics and lipidomics studies aim to measure the variation and co-variation of metabolite or lipid abundances between or across biological sample groups in order to generate new hypotheses for subsequent targeted investigation. Explorative analysis using unsupervised multivariate statistical methods (e.g. principal components analysis, PCA) is carried out first to assess the data reproducibility, detect possible outliers, and visualise possible groupings in the dataset. Subsequently, a univariate statistical test is applied to each peak in the nESI DIMS data to identify specific peaks whose intensities significantly change across different biological sample groups. These tests can be parametric (assuming a normal distribution, e.g. t-test or ANOVA) or non-parametric (e.g. Wilcoxon signed-rank or Kruskal-Wallis). Due to the large number of peaks that are often analysed (and statistically tested), a multiple testing procedure is carried out to reduce the occurrence of peaks incorrectly being identified as significant (i.e. reduce false positives). The optimal approach for DIMS data is to apply the Benjamini-Hochberg false discovery rate (FDR) procedure⁶⁵. A disadvantage of univariate tests is that possible correlations between peaks are not taken into account. Therefore, supervised multivariate approaches (e.g. partial least squares discriminant analysis, PLS-DA) are used as a complementary approach to discover individual metabolites or groups of metabolites that discriminate between different biological sample groups. Note that these models can be used to predict the class (i.e. healthy or disease) of any subsequent biological samples. The prediction accuracy, based on internal cross-validation of the data or an independent test set, is used as a measure of the group separation. Statistical significance of the observed group separation is evaluated with permutation testing. In a permutation test the labels of the biological samples are randomly permuted and a new PLS-DA model is constructed. This process is repeated 1000 (or more) times. Statistical significance of the original PLS-DA model is then assessed by relating the values of the group separation of the permuted data to that of the non-permuted data. Although the permutation test is a powerful statistical procedure, its results should be approached cautiously for very small sample sizes (see experimental design section)^{66,67}. Peaks that contribute to the group separation are identified based on PLS-DA variable importance measures such as the Variable Importance in the Projection (VIP) or the selectivity ratio (SR). To reduce the influence of irrelevant peaks on the model, forward (variable) selection that is based on the VIP or SR can be

carried out before permutation testing.

Metabolite annotation

Automated and robust annotation and identification of hundreds to thousands of peaks in MS-based metabolomics is currently a difficult, complex process and is widely regarded as a bottleneck of data interpretation⁶⁸. Analytical and computational developments are, however, continuing to improve this procedure. Metabolite annotation in spectral stitching DIMS studies starts by assigning elemental compositions to all (or as many as possible) m/z measurements, followed by mapping each elemental composition to a single or multiple chemical name(s). This is done using in-house developed MI-Pack software situated within Galaxy that interacts with public databases (e.g. Human Metabolome Database (HMDB), Kyoto Encyclopedia of Genes and Genomes (KEGG), LIPID MAPS and PubChem)⁶⁹. Due to the finite mass accuracy of mass spectrometers and the complexity of chemical space, this process often results in multiple annotations per peak, however the ultra-high mass accuracy used here minimises the numbers of annotations. To further reduce the number of false positive assignments, we employ several approaches^{69,70}: a single metabolite is often detected as multiple metabolic features, including adducts (e.g. $[M+H]^+$ and $[M+Na]^+$) and naturally occurring isotopes. The latter is often used to reduce the number of incorrect elemental compositions using relative isotopic abundance measurements to calculate the number of atoms (e.g. carbon, sulphur, nitrogen) within the metabolite²⁴. To categorise the confidence of metabolite identification and to make metabolite assignments within a study or across multiple studies comparable, different levels of metabolite annotation and identification have been reported and described by the Metabolomics Standards Initiative (MSI)²⁷. Metabolite assignments in nESI DIMS experiments are based upon m/z values and therefore are reported as putatively annotated (defined as level 2 by the MSI), however the ultra-high mass accuracy and additional steps used in our DIMS workflow (e.g. the use of adduct and isotope patterns) result in a high degree of confidence in the molecular formula(e) annotations. Additional targeted experiments, such as MS^n fragmentation and spectral library matching against an authentic standard, are required to definitively identify the compound(s) of interest (defined as level 1 by MSI).

Data storage and sharing

Metabolomic studies continue to increase in size and produce ever-increasing amounts of experimental data. Open access to research data and knowledge, in a standardised and reproducible way, is important for maximising the value of metabolomics (or any other) datasets. As the requirements of journal publishers (including data journals such as *Scientific Data* and *Gigascience*) and funding bodies to share data and results continue to grow, it is likely and indeed highly preferable that open access to data will become standard practice in metabolomics. This is supported by the increasing number of publicly available repositories for experimental data (e.g. MetaboLights and

Metabolomics Workbench^{71,72}) as well as the number of metabolomics datasets within these public repositories. Recently, a data descriptor for a spectral-stitching nESI DIMS metabolomics study — where the efficacy of a signal-correction algorithm and the reproducibility of a multi-batch study were evaluated — was published in such a data journal and the data is openly available via the MetaboLights repository (MTBLS79)⁷¹. This dataset serves as a benchmark for the metabolomics community, and complements the current publication²⁹.

MATERIALS

REAGENTS

- Pre-collected and frozen (at -80 °C) biological samples: urine, serum, plasma, tissue.
!CAUTION ethical regulations must be followed and patient consent obtained when working with human samples. A licence to undertake animal work should be obtained from the UK Home Office. **!CAUTION** appropriate ethical approval, licensing and training must be sought before working with animals. **!CAUTION** biological samples pose an infection risk and should be handled with adequate personal protective equipment.
- Laboratory collected biological samples: Adherent or suspension mammalian cell cultures, spent media, small whole organisms (e.g. *Daphnia magna* and *Drosophila*).
!CAUTION biological samples pose an infection risk and should be handled with adequate personal protective equipment.
!CAUTION Cell lines should be regularly checked to ensure they are authentic and are not infected with mycoplasma.
- Crushed wet ice.
- Dry ice. **!CAUTION** causes burns and poses an asphyxiation risk. This should be handled with adequate personal protective equipment in a well-ventilated space.
- Liquid nitrogen. **!CAUTION** causes burns and poses an asphyxiation risk. This should be handled with adequate personal protective equipment in a well-ventilated space.
- Nitrogen gas (oxygen free, 30 psi). **!CAUTION** poses an asphyxiation risk. This should be handled in a well-ventilated space.
- HPLC grade chloroform (J.T. Baker, SciChem, 9174). **!CAUTION** chloroform is flammable and toxic and should be handled in a fume hood
- HPLC grade methanol (J.T. Baker, SciChem, 9822) **!CAUTION** methanol is flammable and toxic and should be handled in a fume hood
- HPLC grade water (J.T. Baker, SciChem, 9823)
- Ammonium acetate (99.999%, Sigma-Aldrich, 372331)
- Formic acid (LC-MS Ultra, Sigma-Aldrich, 14265). **!CAUTION** flammable and corrosive; should be handled in a fume hood.
- LTQ Velos ESI Positive Ion Calibration Solution (Thermo Scientific Pierce, 88323). LTQ Velos ESI Negative ion Calibration Solution (Thermo Scientific Pierce, 88324) **!CAUTION** flammable and toxic; should be handled in a fume hood.

EQUIPMENT

- High resolution, high mass accuracy Fourier transform mass spectrometer (Thermo Scientific *LTQ FT Ultra*TM Fourier transform ion cyclotron resonance mass spectrometer; or Thermo

Scientific Q *Exactive*[™] Hybrid Quadrupole Orbitrap[™] mass spectrometer; or Thermo Scientific *Orbitrap Elite*[™] Hybrid Ion Trap-Orbitrap mass spectrometer). All instruments controlled by a dedicated PC running Xcalibur software (Version 2.0 or greater, Thermo Scientific)

- Automated multisample chip-based nESI sample ionisation platform (Advion Biosciences Triversa Nanomate) coupled to the mass spectrometer. This is controlled by ChipSoft software (Version 8)
- nESI 400-nozzle electrospray chip (Advion Biosciences, 1003446)
- nESI electrospray 400 tip rack (Advion Biosciences, 1004763)
- Glass Duran bottles (50 mL and 100 mL, Fisher Scientific)
- Two Hamilton syringes (750SNR 500 µL [Thermo Fisher, 80865])
- High-quality solvent resistant pipette tips (Fisherbrand, Thermo Fisher. 200 µL: 11587442, 1000 µL: 10787524)
- High-quality solvent resistant plastic microfuge tubes (Eppendorf. 1.5mL: 10509691, 2 mL: 10038760)
- 1.75 mL tall form glass vials with aluminium-lined screw caps (Wheaton, 151061)
- Glass Pasteur pipettes (Volac, D810, 150 mm)
- Precellys®24 bead-based homogeniser (Bertin Technologies)
- 2mL tubes for Precellys®24 homogeniser (2mL tubes with 1.4 mm diameter ceramic beads [Stretton Scientific, 03961-1-003])
- Vortexer mixer
- Refrigerated Centrifuge (Microcentrifuge Biofuge Primo R [Thermo Scientific, 75005440]; Bucket rotor [Thermo Scientific, 75007591]; Fixed angle rotor [Thermo Scientific, 75007593])
- Fume hood
- Nitrogen blow down drier (Techne, Fisher Scientific. Sample concentrator [11593859], Heating block [11563799], 2x Dry blocks [11523809], Needles [11426578])
- SpeedVac concentrator (Vapour trap [Thermo Scientific, RVT5105230]; SpeedVac [Thermo Scientific, SPD111V230]; Pump [Knf, LABOPORT N820(.3) FT.18])
- Carbon tips for loading non-polar (lipid) samples into well plates (Advion Biosciences, 1004763)
- Plastic well plates (Fisher Scientific: 96-well [AB-0800]; 384-well [TF-0384])
- Self-adhesive aluminium sealing tape (Corning, 6570)
- Heat sealing plate covers (Fisher Scientific, AB-1720)
- Well plate heat sealer (Thermo Scientific, ALPS 50 V)
- Refrigerator (4 °C)
- Freezers (-20 °C and -80 °C)
- Xcalibur software (Thermo Fisher, version ≥ 2.0)
- An instance of Galaxy and Galaxy-M tools, workflows and associated software (see <https://github.com/Viant-Metabolomics/Galaxy-M> for a detailed installation guide)

- A PC (Microsoft Windows 7, Intel Core i7, 2.8 GHz, 4.00 GB Ram) with internet access

REAGENT SETUP

Biological sample collection and handling To halt enzymatic activity and ensure that the biological sample is a true representation of the phenotype being tested, samples (including tissue or biofluid, e.g. plasma, serum, urine) should be frozen as soon as possible after sampling. Small whole organisms, e.g. *Daphnia*, *Drosophila*, should be placed in a microfuge tube and flash-frozen in liquid nitrogen. Adherent or non-adherent cultured mammalian cells should be quenched using -40 °C 60% v/v HPLC grade methanol (on dry ice) as the quenching solution following existing protocols^{45,73}. To enable the calculation of cell biomass, the quenching tube is weighed at the start (prior to addition of any quenching solvents) and at the end (when only frozen cells are present). **CRITICAL** All biological samples must remain frozen at -80 °C until analysis. The minimum recommended sample sizes for this high sensitivity method are 1 mg biomass of tissue, mammalian cells or whole organism, or 5 µL of biofluid.

!CAUTION Ethical approval must be sought and ethical protocols followed when working with human samples.

Solvent Preparations Only use high quality plasticware (tips and tubes as recommended above) to minimise the leeching of plasticisers into the extraction solutions. Ensure that each set of consumables (tips and tubes) and solvents are from the same lot number. Extraction solvents for tissue and mammalian cells (methanol, chloroform and water, each on their own) and biofluids (methanol:chloroform in a 1:1 solution and water on its own) should be prepared and decanted into solvent-rinsed and dried 50 or 100 mL Duran bottles. Methanol:chloroform 1:1 should be pre-cooled to -20 °C for >2 hr. All other solvents should be pre-chilled on wet ice for at least 30 min. Separate vials of chloroform and methanol should be prepared to wash the Hamilton syringe that is used to remove metabolic or lipid extracts (see Steps 1A xi – xii and 1B v). DIMS analysis solvents should be prepared as follows: (i) positive ion polar solvent – 4:1 (v/v) methanol:water with a total of 0.25% (v/v) formic acid; (ii) negative ion polar solvent – 4:1 (v/v) methanol:100 mM aqueous ammonium acetate; (iii) non-polar solvent (suitable for positive and negative ionisation analyses) – 2:1 (v/v) 7.5 mM methanolic ammonium acetate:chloroform (dissolve the ammonium acetate in the methanol prior to chloroform addition). All solvent mixtures are stable at room temperature for at least one week.

EQUIPMENT SETUP

Equipment used during extraction Pre-cool both the swinging bucket and fixed-angle microtube centrifuges to 4 °C. Turn on Precellys®24 homogeniser to allow it to initiate. Turn on the cold trap that is connected to the SpeedVac concentrator before starting the extraction procedure (requires approximately 2 hr to achieve operating temperature).

BOX 1: Triversa and mass spectrometer setup. Timing 1 h

(A) Initial system set-up

1. If the mass spectrometer has been in standby mode, allow at least 2 hr to equilibrate once fully switched on.
2. Open the MS acquisition viewer on the PC (MS Tune for Thermo systems). Calibrate the MS according to the manufacturer's instructions using the calibration mixture.
3. Attach the Triversa onto the bracket on the front of the MS.
4. Load the 400-nozzle nESI chip, 400 tip set and a 96-well plate containing a test biological sample (e.g. biological extract in MS solvent, Table 2).
5. Ensure that the nitrogen cylinder (or generator) supplying the Triversa has sufficient gas remaining (recommended to have >1000 psi for a 24 hr run) and has sufficient backpressure.
6. Position the Triversa so that it is central to the MS source and 3-5 mm away from the source by adjusting the positional settings in the ChipSoft software (*Method Manager / Spray Optimisation*).

(B) Triversa method creation, electrospray optimisation and mass spectrometry tuning to enhance detection sensitivity

1. Create a method file with the recommended settings:
 - (a) Sample volume **7 µL (polar), 10 µL (non-polar)**
 - (b) Vent Headspace **ON**
 - (c) Aspirate air after sample **ON**
 - (d) Volume of air to aspirate after sample **0.2 µL (polar), 2.0 µL (non-polar)**
 - (e) Trigger acquisition when Input Signal received **ON**
 - (f) Spray sensing **OFF**
 - (g) Gas pressure **0.3 psi**
 - (h) Voltage to apply **1.5kV**
 - (i) Positive ion or Negative ion **as desired**
2. To optimise the nanoelectrospray and DIMS detection for each specific biological sample type inject a test sample by using the 'Spray Optimization' tool (within the 'Method Manager' tab). The Triversa electrospray current should be stable and between 100-300 nA (for problems see

?TROUBLESHOOTING Table 3). **CRITICAL** A stable electrospray current is required to generate reproducible data. Electrospray stability is partially dependent on the voltage and gas pressure settings on the Triversa (recommended settings above). These settings are sample type dependent and therefore may require optimisation to improve electrospray stability.

3. Optimise the Triversa voltage: set the gas pressure to 0.3 psi and the voltage to 1.5 kV, then—while monitoring the spray stability on the mass spectrometer—decrease the voltage in steps of 0.1 kV until the spray is lost. Switch the polarity mode on the Triversa then switch back to the polarity that was originally being used. If the spray does not return add 0.2 kV to the current voltage setting and this is now the optimal voltage setting that should be used.

4. Optimise the Triversa pressure: set the voltage to its optimised setting and the pressure to 0.3 psi. Then alter the pressure in steps of 0.05 psi and observe any improvement or deterioration of the electrospray current.

5. Optimise the position of the Triversa in relation to the MS source: observe the ion intensity and reproducibility in the mass spectrometer using the mass spectrometer detector program (MS Tune). Move the Triversa closer or further away from the mass spectrometry source (by changing settings in *Method Manager / Spray Optimisation*). To increase the sensitivity move the Triversa closer to the MS, to increase reproducibility move it further away (the optimal distance should be in the range 3-5 mm). For problems with DIMS detection see **?TROUBLESHOOTING** Table 3.

6. Where available, it is recommended to optimise peak detection sensitivity by using the automated tuning function (within the mass spectrometer detector program, MS Tune) on ions within the mass ranges of interest, e.g. ions around 75 m/z ('low tune'), 150 m/z ('medium tune') and 400 m/z ('high tune') for polar extracts. Each m/z range optimisation is carried out independently by repeating the tuning at that m/z until no further increase to signal intensity is achieved, then saving the optimal tune parameters as separate tune files (i.e. 'low tune', 'medium tune', 'high tune').

(C) Creation of the spectral-stitching method

1. Using the *Instrument Setup* tab in Xcalibur software, create a method with multiple segments (for the case where transient scans are collected — LTQ FTTM and LTQ FT UltraTM) or multiple scan events (for case where transient scans are not available — Q ExactiveTM and Orbitrap EliteTM) that represent the m/z windows for the spectral-stitching method (widths, acquisition length, resolution and AGC detailed in Figure 2 and Table 1; see Supplementary Methods for Q ExactiveTM and Orbitrap EliteTM instrument method files). When using scan events, set the number of acquired microscans to 10 — this averages 10 transient scans.

2. If available, assign a tune file to each window: 'low tune' for windows with a centre <100 m/z ; 'medium tune' for windows with a centre 100-250 m/z ; 'high tune' for windows with a centre >250 m/z . Otherwise assign an appropriate tune file for the entire mass range (e.g. optimise at the mid- m/z range being detected or within the specific m/z range of interest).

3. At the beginning of the method apply a 30 s delay before MS data is collected in order to allow stabilisation of the nESI spray current.

[BOX END]

Establishing Galaxy-M-workflows Data processing and analysis steps described here are collated into Galaxy specific tools and workflows, referred to as Galaxy-M²⁸. An extensive guide on how to install Galaxy-M is available from GitHub (<https://github.com/Viant-Metabolomics/Galaxy-M>). Here, tools and workflows within Galaxy-M are well documented, including default parameters, and have self-explanatory names to guide the user through the data processing and analysis steps. Additional reading and tutorials on how to use a Galaxy instance are available online (<https://wiki.galaxyproject.org/Learn>).

Most DIMS studies include a large number of files and a considerable amount of data. Galaxy-M assumes that the user will store the data (e.g. .RAW files with or without transient data) on a file system that is directly accessible by the Galaxy instance (i.e. no Galaxy-upload required).

PROCEDURE

Metabolite and lipid extraction from biological samples **TIMING 2-4 hr per batch**

CRITICAL To minimise inconsistencies across the handling of biological samples and therefore to reduce sample-to-sample variation, extract no more than 20 samples in a single batch. Repeat this section until all samples are extracted. **CRITICAL** Randomise the extraction order to ensure that biological sample classes are randomised within and across extraction batches. **CRITICAL** An *extract blank* sample must be prepared. Here, the extraction procedure is carried out in the absence of biological sample, and the resulting *extract blank* is used to identify compounds that arise solely from sample preparation procedures.

1| Perform the steps in Option A if you are extracting from tissue whole organisms or mammalian cells. Perform the steps in Option B if you are extracting from biological fluids.

A Extraction from tissue, whole organisms or mammalian cells

- i. Weigh the biological samples ensuring that they do not thaw by keeping them on dry ice. Mammalian cell biomass should be estimated during the quenching step (see *REAGENT SETUP - Biological sample collection and handling*, above).
- ii. For tissues or whole organisms place the biological material into a labelled Precellys®24 tube and set on dry ice.. For mammalian cells place the biological material into a labelled high-quality 2 mL microfuge tube and set on dry ice.
- iii. Add 8 µL of ice-cold methanol for every mg of frozen biological sample mass (8 µL / mg) (using high-quality solvent resistant plastic pipette tips).
- iv. For tissues or whole organisms, homogenise the samples using a Precellys®24 system (2 × 10 s bursts of 6400 rpm separated by a 5 s gap) before returning the tubes to dry ice. For mammalian cells, vortex samples (30 s) before returning the tubes to dry ice.

- v. Label a 1.75 mL glass vial per sample and add 8 μ L / mg ice cold chloroform (using a clean glass Hamilton syringe) to each vial and set on wet ice.
- vi. Using a glass Pasteur pipette, transfer the homogenate from either the Precellys®24 tube (tissue or whole organism samples) or microfuge tube (cultured cells samples) into the 1.75 mL glass vial containing chloroform. Vortex each vial (15 s) and then return to the wet ice.
- vii. Add 7.2 μ L / mg HPLC-grade water to each vial, vortex (30 s) then return to the ice (samples should be cloudy in appearance).
- viii. Leave samples on wet ice for 10 min to allow metabolites and lipids to partition between the polar and non-polar solvents.
- ix. Centrifuge samples (2500-g, 4 °C, 10 min) using a swinging bucket centrifuge to induce phase separation. **CRITICAL STEP** remove from centrifuge bucket extremely carefully to avoid disturbing the protein interface and biphasic separation of solvents.
- x. Set the samples on the bench at room temperature for 5 min to allow completion of the phase separation. Sample is now bi-phasic, with the polar (upper) and non-polar (lower) layers separated by an interface of denatured proteins and cell debris.
- xi. Remove a fixed volume of the polar phase into a microfuge tube using a clean Hamilton syringe (typically 300 μ L, with the option to collect multiple aliquots). Rinse the syringe twice with methanol wash solvent (2 \times 500 μ L) between each biological sample.
- xii. Remove a fixed volume of the non-polar phase into a new 1.75 mL glass vial tube using a clean Hamilton syringe (typically 150 μ L, with the option to collect multiple aliquots). Rinse the syringe twice with chloroform wash solvent (2 \times 500 μ L) between each biological sample. To obtain the non-polar phase slide the Hamilton syringe down the side of the glass vial and gently move the protein interface out of the way with the syringe. **CRITICAL STEP** removing the non-polar layer requires care so as to avoid debris attaching to the Hamilton syringe.
- xiii. Dry the polar extracts in a SpeedVac concentrator using no heat (1-2 hr). **!CAUTION** Ensure the associated cold trap has reached operating temperature (< -100 °C) to prevent solvents being vented to the atmosphere.
- xiv. Dry non-polar biological samples under a stream of nitrogen (15 min) and cap the vials quickly to ensure a nitrogen-rich atmosphere remains inside.. **<CRITICAL STEP>** This removes oxygen from the sample to prevent lipid oxidation and peroxidation. **!CAUTION** This must be done in a fume hood to prevent the escape of chloroform into the laboratory.
- xv. Store dried extracts frozen at -80 °C or proceed to step 2. **PAUSE POINT** extracted biological samples can be stored at -80 °C for up to a few months.

B *Extraction from biofluids (serum, plasma, urine)*

- i. Thaw biofluid on wet ice for 30-60 min, then vortex (15 s) and remove a fixed volume (e.g. 200 μ L) into a 2 mL high-quality solvent resistant plastic microfuge tube. Keep samples on wet ice.
- ii. Add 600 μ L methanol:chloroform (1:1, -20 °C), vortex (15 s) and return sample to the ice.
- iii. Add 300 μ L water, vortex (15 s) and return samples to the ice.

- iv. Centrifuge samples (20,000-g, 4 °C, 5 min) to induce phase separation.
- v. Follow steps 1A x-xiv (above) except take a 350 µL aliquot from the polar phase and a 100 µL aliquot from the non-polar phase.
- vi. Store dried extracts frozen at -80 °C or proceed to Step 2. **PAUSE POINT** extracted biological samples can be stored at -80 °C for up to a few months.

Resuspension of biological sample extracts in DIMS solvent **TIMING 1-2 hr**

CRITICAL STEP Prior to biological sample resuspension, it is advisable to create and analyse a serial dilution of resuspended extracts to establish the optimal concentration that maximises peak counts (of biological origin) and stability of the nESI current for the specific biological samples being analysed. Recommended starting points for the resuspension solvents (which include sample modifiers, formic acid and ammonium acetate, to enhance electrospray efficiency) and volumes are detailed in Table 2. This is carried out using a spare biological extract which will be used up and not form part of the subsequent metabolomics analyses.

- 2 | Add the appropriate type and volume of DIMS solvent to the extracts as described in Table 2 (or the volume determined in the method optimisation experiment) and vortex (30 s) to dissolve. **CRITICAL STEP** lipid concentrations should not exceed 10 pmol/µL to prevent the formation of lipid aggregates³⁴.
- 3 | Create a QC sample by taking a fixed volume from each biological sample and pooling, to a minimum total volume of 180 µL (enough for 6 × QC analyses). As the number of biological samples analysed increases, so too should the minimum QC volume, such that a QC sample can be analysed every 4-6 biological samples throughout an analytical run. **CRITICAL STEP** If data will be collected in several batches, an identical QC sample must be used across all batches. This will allow successful application of the signal correction algorithm (Step 28).
- 4 | Centrifuge samples (polar extracts in microfuge tubes: fixed angle rotor, 20,000-g, 4 °C, 10 min; non-polar extracts in glass vials: bucket rotor, 2,500-g, 4 °C, 10 min) to remove particulates that can adversely affect nESI stability. **CRITICAL STEP** Particulates in the sample can fully or partially block the nESI nozzle leading to loss or instability of electrospray.
- 5 | When loading samples, place a 96-well or 384-well plate on wet ice (to prevent sample evaporation) and pipette 3 × 10 µL of each sample into triplicate (consecutive) wells along a row using either Fisherbrand tips (polar) or carbon tips (non-polar). On completion of each well plate row, cover with self-adhesive aluminium sealing tape to prevent solvent evaporation. **CRITICAL STEP** To avoid particulates in the sample wells, keep the samples at the same fixed angle as they were in the centrifuge and avoid disrupting the pellet (likely invisible to human eye) when pipetting. **CRITICAL STEP** Ensure that the biological sample phenotypes are randomised in the well plate (facilitating a randomisation of the analytical run order). Ideally, the sample order should start with a series of trial samples, typically MS diluent (×3 technical replicates) and QC samples (×3 technical replicates), which allow the user to assess

the quality of the electrospray and DIMS detection. This is followed by the *extract blank* sample (×3 technical replicates) and by a QC sample (×3), and then the biological samples (each ×3) with a QC sample (×3) every 4-6 samples, and finally end with a QC sample (×3) (see Figure 3 for an example). The analysis of QC samples at regular intervals throughout the sample sequence run is essential to enable *signal correction* data processing (Step 28).

- 6 | Once completed, remove the self-adhesive aluminium sealing tape and heat-seal the entire well plate with the appropriate foil (170 °C for 2 seconds).
- 7 | Open ChipSoft software (controlling the Triversa) and in *Interface Settings* set the plate cooler to 10 °C. Put the well plate on to the Triversa and allow at least 30 min for the samples to equilibrate to 10 °C. **CRITICAL STEP** variations in sample temperature can alter the electrospray process leading to poor technical reproducibility

DIMS analysis TIMING ~9 min per polar biological sample (comprising of triplicate technical replicate analysis on the Q Exactive™, Orbitrap Elite™ or LTQ FT Ultra™; including sample loading and equilibration)

- 8 | Ensure the well plate is located on the Triversa and cooled to 10 °C. Load a new electrospray chip and 384-tip tray into the Triversa and empty the tip waste tray. Align the Triversa with the MS source (Box 1). Within the ChipSoft software (in *Sequence View*) create a sample sequence using the optimal Triversa method (Box 1). **CRITICAL STEP** Ensure that the Triversa contact closure is set to '*Trigger acquisition when Input Signal received*', which enables the Triversa to initiate MS data collection when the nESI begins.
- 9 | Within the MS vendor software (Xcalibur for Thermo instruments) create a sample run order using the method created in Box 1. **CRITICAL STEP** For FT-ICR instruments, ensure that the collection of time domain (transient) data is enabled (Within the Tune program: Diagnostics → Toggles → Include FT transient).
- 10 | Initiate the Triversa sample run and when the *ChipSoft Virtual Device* window states that it is awaiting input from the MS system, initiate the MS sample run in Xcalibur. . **PAUSE POINT** DIMS data can be stored for unlimited duration until data processing is initiated.

Data processing: spectral-stitching and signal filtering TIMING 6-12 hr dependent on the number of biological samples

CRITICAL STEP At any point in the data analysis process, the Galaxy history can be downloaded, archived and shared with others.

- 11 | Create a text file (as comma separated value [.csv] format, e.g. *samples.csv*) to function as a sample identifier for the samples being processed. This contains four columns separated with a comma: first column – spectral file names, second column – class identification (e.g. control, QC, etc.), third column – batch number and fourth column – run order.

- 12 | Open an internet browser and go to the Galaxy-M homepage (e.g. <http://localhost:8080> for a local Galaxy instance, see *Establishing Galaxy-M-workflows* in *EQUIPMENT SETUP*, above).
- 13 | Select **Tools** → **Get Data** → **Upload Data** from the left panel. This brings up the upload data form. Upload the .csv file created in Step 11 (ensure “.csv” is selected as the file type).
- 14 | Use **Tools** → **Get Data** → **File List Manager** (left panel) to create a file list which specifies the location of the files on the user’s PC and the type of MS instrument used to collect the data.
- 15 | If the DIMS data includes transient files, proceed to Step 16. If no transient data is available then proceed to Step 18. <CRITICAL STEP> Transient data is the free induction decay mass spectrometry data, which requires processing as in Steps 16-17. For data in the processed vendor format (*.RAW), these processing steps have already been completed by the vendor instrument software.
- 16 | Run **Tools** → **SIM-stitching** → **Sum Transients**. This sums the transient data for the mass spectral SIM windows within each technical replicate. To avoid the inclusion of low quality transients (caused by poor or unstable nESI), a total ion current (TIC) threshold can be set (optional). Transients are removed according this threshold.
- 17 | Run **Tools** → **SIM-stitching** → **Process Transients**. This applies Hanning apodisation, zero-filling, Fast Fourier Transform and baseline correction to the summed transients.
- 18 | [Optional] Create a tab-separated text file (.txt format) that will enable internal mass calibration of the spectra. This file contains the names (in the first column) and the exact mass (in column 2) of compounds known to be present in the sample (e.g. *calibrants.txt*). If this information is not known, this step can be omitted and data will be externally calibrated (almost certainly resulting in lower mass accuracy). **CRITICAL STEP** To create a true calibrant list additional MS experiments (e.g. fragmentation) are often needed to annotate (i.e. elemental composition) and/or identify (i.e. structure) the m/z values. Compounds detected in the extract blank sample can be used for calibration across different MS studies when identical extraction protocols and instruments have been used.
- 19 | Run **Tools** → **SIM-stitching** → **Mass Calibration and SIM-stitching**. This stitches the multiple m/z windows together (removing any m/z overlap in the process) and internally calibrates the mass-to-charge axis of the spectra (optional). The required noise threshold, as a signal-to-noise ratio (SNR), must be stipulated (typically 3:1; data with intensities below this threshold will be discarded). If internal calibration is required, provide *calibrants.txt* (Step 18) and specify the minimum intensity (typically >50 SNR) and the maximum mass tolerance (typically 2.0 ppm) of peaks to be used for internal calibration.
- 20 | Run **Tools** → **SIM-stitching** → **Replicate Filter**. This step retains peaks present within at least 2 of the triplicate measurements of each biological sample (and averages their intensities) and thus functions to remove noise features. The maximum m/z ppm range a peak must fall within (across the triplicate measurements) for it to be counted as the same peak must be stipulated (typically 1.5 ppm; **CRITICAL STEP** If the mass accuracy of the data

is less than optimal this m/z tolerance can be increased to ensure that real peaks are retained).

- 21 | Run **Tools** → **SIM-stitching** → **Align Samples**. This aligns the peaks across all of the *Replicate Filtered* biological samples. The maximum m/z ppm range a peak must fall within (across the biological samples) for it to be counted as the same peak must be stipulated (typically 2.5 ppm; **CRITICAL STEP** If the mass accuracy of the data is less than optimal this m/z tolerance can be increased to ensure that sample-related peaks are retained).
- 22 | Run **Tools** → **SIM-stitching** → **Create DSO**. This converts the DIMS data into a DataSet Object (DSO – a construct of Eigenvector Research to facilitate data set handling and sharing [<http://www.eigenvector.com/software/dataset.htm>]) .xml file, which includes a peak intensity data matrix, row labels (sample classes), column labels (m/z values) and information provided in the text file in Step 11. The data within the DSO is updated with each additional processing step.
- 23 | Run **Tools** → **SIM-stitching** → **Blank filter**. This compares biological sample peaks to those appearing in the *extract blank sample*. Peaks that appear in both are discarded according to the defined biological:blank peak intensity ratio (typically peaks are retained if this ratio is >3).
- 24 | Run **Tools** → **SIM-stitching** → **Sample Filter**. This retains biological peaks that are present within a defined minimum percentage of all biological samples (80% for initial biological interpretation; lower if the experimental design suggests this; see *data procession* section in the *INTRODUCTION*, above).
- 25 | Run **Tools** → **SIM-stitching** → **Missing Value Sample Filter**. This calculates the number of missing values contained in each sample within the DSO data matrix. Samples with unacceptably high numbers of missing values can be automatically excluded.

Normalisation, signal correction, data quality assessment, missing value imputation and generalised log (G-log) transformation **TIMING** 30 min – 1 hr

- 26 | Run **Tools** → **Matrix Processing** → **PQN Normalisation**. This normalises the data matrix (DSO) from Step 25. The normalisation quotients for this are calculated from the DSO in Step 22 using peaks that are present in all samples.
- 27 | [Optional] Run **Tools** → **Matrix Processing** → **Peak Outlier Detection**. This tool detects outliers detrimental to the effectiveness of the signal intensity correction in **step 28**. For each peak, it fits a 2nd or 3rd order polynomial regression curve to the non-QC peaks. Only non-QC data points are used in this curve fitting to avoid statistical bias. A confidence interval for each curve is calculated (typically 95 or 99%) and any (QC) data point lying outside the confidence interval is considered an outlier, and can be removed⁶¹.

- 28 | [Optional] For single or multiple batch studies, the signal intensity correction tool can be applied. Run **Tools** → **Matrix Processing** → **Signal correction and Peak Quality Filtering**. **For each peak**, the tool applies robust cubic smoothing splines to the QC data points ordered by injection. The resulting fitting curves are then used to correct for the technical variation arising from inter-batch measurements⁶¹. This approach can also be used to correct signal-intensity technical variation that arises within a single batch. Additionally, peaks that are highly variable (e.g. RSD >20%) across batches can be filtered based on different criteria (e.g. RSD).
- 29 | Run **Tools** → **Matrix Processing** → **Missing Value Imputation**. This imputes any missing values that are present in the normalised data matrix using the K-nearest neighbour method. K can be set by the user (default is 5).
- 30 | [This step is only required prior to multivariate statistics] Run **Tools** → **Matrix Processing** → **G-log Transformation**. This optimises the lambda parameter from the QC data matrix (i.e. QC samples within the data matrix are automatically selected), then applies the generalised log transformation to the normalised data matrix.

Statistics Analyses **TIMING** 1 – 2 hr

- 31 | Univariate statistical analysis can be applied to the normalised data matrix immediately before missing value imputation (from Step 28) or after missing value imputation (from Step 29). Run **Tools** → **Statistics** → **Univariate Analysis**, or export the data for further analysis using external statistical packages (**Tools** → **DataSet Object**).
- 32 | Multivariate statistical analysis can be applied to the generalised log transformed normalised data matrix (from Step 30). Run **Tools** → **Statistics** → **PCA and Scores Test**, or export the data for further analysis using external statistical packages (**Tools** → **DataSet Object**).

Metabolite Annotation **TIMING** 1 hr, and significantly longer (>8 hrs) when Molecular formulae Search is used

- 33 | Run **Tools** → **DataSet Object** → **Get Peak List** to create a peak list from the biological sample data matrix (Step 25).
- 34 | [Optional step; time-consuming] Run **Tools** → **MI-Pack** → **Molecular formulae Search**. This matches each peak m/z value to one or more elemental compositions ($C_cH_hN_nO_oP_pS_s$, including adducts, such as $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$) within a mass error tolerance stipulated by the user (e.g. typically within ± 1 ppm for polar or ± 2 ppm for non-polar FT-ICR MS analysis).
- 35 | Run **Tools** → **MI-Pack** → **Single-Peak Search**. This assigns a putative biological compound identification to each m/z value (or elemental composition) within a mass error tolerance

stipulated by the user (typically within ± 1 ppm for FT-ICR MS) using database(s) specified by the user (e.g. KEGG, HMDB, LIPID MAPS).

36 | Run **Tools** → **MI-Pack** → **Peak-Pattern Search**. This identifies adduct/isotope patterns (e.g. ^{12}C - ^{13}C and $[\text{M}+^{39}\text{K}]^+ - [\text{M}+^{41}\text{K}]^+$) and relative isotopic abundance measurements within the data.

37 | Run **Tools** → **MI Pack** → **Combine Outputs**. This produces a summary file of the outputs produced by Steps 33-36.

Public repositories for metabolomics studies

38 | We recommend that all the DIMS data, including .RAW files and data matrices, should be deposited into a public repository (i.e. MetaboLights⁷¹ [<http://www.ebi.ac.uk/metabolights>] and/or Metabolomics Workbench⁷² [www.metabolomicsworkbench.org]). To do this, follow the instruction on the data repository's websites. A stable identifier will be assigned to the dataset, which can then be used to cite the dataset in a publication.

TROUBLESHOOTING

See Table 3 for troubleshooting guidelines.

TIMING

Step 1A-1B, Metabolite and lipid extraction from biological sample: 2-4 hr per batch

Step 2-7, Resuspension of biological sample extracts in DIMS solvent: 1-2 hr

Step 8-10, DIMS analysis: ~9 min per polar biological sample (comprising of triplicate technical replicate analysis on the LTQ FT UltraTM; Q ExactiveTM or Orbitrap EliteTM including sample loading and equilibration)

Step 11-25, Data processing: spectral-stitching and signal filtering: 6-12 hr dependent on the number of biological samples

Step 26-30, Normalisation, batch correction, data quality assessment, missing value imputation and generalised log (G-log) transformation: 30 min – 1 hr

Step 31-32, Statistics Analyses: 1 – 2 hr

Step 33-37, Metabolite Annotation: 1 hr, and significantly longer (>8 hrs) when Molecular formulae Search is used

ANTICIPATED RESULTS

The spectral-stitching nESI metabolomics / lipidomics method — in which data is collected as a series of overlapping m/z windows that are ‘stitched’ together (Figure 2) — generates high quality data in less than 10 min per biological sample (including sample loading and equilibration, and triplicate sample acquisition; Table 1). This approach significantly increases the dynamic range of the spectrum (Figure 4)^{23,24} without increasing the number of ions entering the detector, which could lead to space-charge effects and poor mass accuracy⁵⁶. The number of detected peaks increased >5-fold compared to a single wide-scan m/z collection (Figure 4)^{23,24}. The use of nESI is critical because it lowers ionisation suppression compared to standard ESI, which also enhances peak detection²⁵. Detection sensitivity is further increased by applying the method on an instrument with a physically larger detector that allows more ions to enter without increasing space-charge effects: a 3-fold increase in peak detection was observed on the LTQ FT UltraTM, with larger detector, compared to the LTQ FT²⁴. For each type and size of MS detector, the spectral-stitching method including m/z window width should be optimised to maximise the number of peaks detected (Table 1).

The mass accuracy of data generated by the spectral-stitching nESI method is ± 1 ppm for peaks >600 m/z ^{23,24}. Collecting the data as multiple m/z windows aids the internal mass calibration because each window can be m/z shifted independently of the others. The ultra-high mass accuracy achieved by this method enables molecular formula(e) annotation of peaks and allows the assignment of putative metabolite and lipid names. Expected numbers of detected and annotated metabolite and/or lipid features vary dependent on sample type. Negative ion spectral-stitching nESI metabolomics of ca. 1 mg of *Daphnia pulex-pulicaria* detected and putatively annotated 1973 and 369 peaks (± 1.5 ppm mass tolerance), respectively, while negative ion lipidomics of ca. 10^7 (ca 10 mg) human leukaemia cells detected and putatively annotated 2934 and 898 peaks (± 2 ppm tolerance), respectively. A lipid spectrum in which selected peaks have been annotated is shown in Figure 5.

The spectral-stitching nESI DIMS method demonstrated high analytical reproducibility: median peak intensity RSD measurements across the QC samples were about 8% for both the LTQ FTTM and LTQ FT UltraTM, about 13.5% for the (LTQ) Orbitrap EliteTM and about 16% for the Q ExactiveTM (Table 4). Based upon these results, it suggests that nESI DIMS utilising a hybrid instrument with a linear ion trap (i.e. LTQ) yields more reproducible peak intensities. These levels of reproducibility have been shown to be maintained in large-scale studies in which data was collected over several days. For example, PCA of a large *Daphnia pulex-pulicaria* metabolomics study had highly reproducible QC samples that were tightly clustered on the PCA scores plot (Figure 6)⁷⁴. Such low technical variance enabled biological variance to be visualised, with a clear distinction of the metabolomes of F0 and F1 *Daphnia pulex-pulicaria* generations (Figure 6)⁷⁴. When the spectral-stitching nESI DIMS method was used for lipidomics, reliable and meaningful biological information to identify drug-induced lipid changes in leukaemia cells was generated⁷. Metabolite intensity measurements were shown to be robust and comparable to those made by NMR spectroscopy, i.e. tyrosine, creatine and phosphocreatine intensity measurements derived from the same sample by the two analytical methods were strongly correlated (Figure 7)¹¹. A more extensive comparison of the spectral-stitching

nESI DIMS and NMR approaches recently showed that the DIMS approach is capable of relative quantification comparable to NMR for several metabolites³⁰. NMR spectroscopy is a quantitative technique that is well-established in the field of metabolomics, which in turn demonstrates the robustness of the spectral-stitching method.

To conclude, the spectral-stitching nESI DIMS method is reproducible, sensitive and high-throughput with ultra-high mass accuracy. These attributes make it an ideal tool for rapid and large-scale metabolomics or lipidomics analyses to identify phenotype-induced perturbations to the metabolome or lipidome.

ACKNOWLEDGMENTS A.D.S. was funded by a Bloodwise programme grant awarded to FL Khanim and MT Drayson. R.J.M.W., M.R.V., J.E. and M.R.J. were funded by the UK Natural Environment Research Council (NE/I008314/1, NE/J017442/1, NE/K011294/1 and R8-H10-61) and R.J.M.W. and M.R.V. by the UK Biotechnology and Biological Sciences Research Council (BB/M019985/1). M.R.J. wishes to also thank Thermo Fisher Scientific for CASE funding. The LTQ FT UltraTM used in this research was obtained through the Birmingham Science City Translational Medicine: Experimental Medicine Network of Excellence project, with support from Advantage West Midlands (AWM). We thank Nadine Taylor (University of Birmingham, UK) and collaborators (John Gunn; Laurentian University, Canada, and Jim McGeer; Wilfrid Laurier University, Canada) who kindly provided us with experimental data for Figure 6. We thank David Broadhurst (University of Alberta) for on-going development and support for the batch correction algorithm.

AUTHOR CONTRIBUTIONS

A.D.S. and **R.J.M.W** co-developed the analytical methods and computational workflow, acquired data and wrote the manuscript. **J.E.** contributed to developing the computational workflow and writing the manuscript. **M.R.J.** contributed to developing the analytical methods and acquired data. **M.R.V.** co-developed the analytical methods and computational workflows, edited the manuscript, and was the academic lead.

COMPETING FINIANCIAL INTERESTS The authors declare that they have no competing financial interests.

Tables

Table 1: The optimised parameters for the spectral-stitching nESI DIMS metabolomics and lipidomics methods. Selected ion monitoring (SIM) mode is used throughout. For a schematic of the spectral-stitching nESI DIMS method see Figure 2.

	LTQ FT	LTQ FT Ultra TM		Q Exactive ^{TM a}		Orbitrap Elite ^{TM a}	
Extract type	POLAR	POLAR	NON-POLAR	POLAR	NON-POLAR	POLAR	NON-POLAR
<i>m/z</i> range covered	70–500 <i>m/z</i>	70–590 <i>m/z</i>	70–1010 <i>m/z</i>	50–620 <i>m/z</i>	50–1020 <i>m/z</i>	50–620 <i>m/z</i>	190–1200 <i>m/z</i>
Total number of <i>m/z</i> windows	21	7	13	14	24	10	18
SIM window width	30 <i>m/z</i>	100 <i>m/z</i>		50 <i>m/z</i>		75 <i>m/z</i>	
<i>m/z</i> window acquisition time	15 s (~10 transients)	15 s (~10 transients)		~5.5 s (minimum of 10 micro scans)		~7.8 s (minimum of 10 micro scans)	
Overlap between <i>m/z</i> windows	10 <i>m/z</i>	30 <i>m/z</i>		10 <i>m/z</i>		20 <i>m/z</i>	
AGC target	1 × 10 ⁵	1 × 10 ⁶		5 × 10 ⁵		5 × 10 ⁵	
Peak resolution	100k for ion at 400 <i>m/z</i>	100k for ion at 400 <i>m/z</i>		140k for ion at 200 <i>m/z</i>		240k for ion at 400 <i>m/z</i>	
Acquisition	5 min 45 s ^b	2 min 15 s ^b	3 min 45 s ^b	2 min 30 s ^b	3 min 30 s ^b	1 min 50 s ^b	2 min 50 s ^b

time per injection							
-----------------------	--	--	--	--	--	--	--

^aSee Supplementary Method for Q Exactive™ and Orbitrap Elite™ instrument method files

^bIncluding an initial 30 s start delay of blank or dummy scans to allow nESI to stabilise.

Table 2: Directions for the resuspension of polar and non-polar extracts for nESI DIMS analysis. Formic acid and ammonium acetate are sample modifiers that enhance electrospray efficiency.

Sample	Ion mode	Resuspension solvent	Recommended resuspension concentration
Tissue, whole organism or mammalian cell polar extract	Positive ion	4:1 v/v methanol:H ₂ O. 0.25% (v/v) formic acid added.	Half of the original extract volume
	Negative ion	4:1 v/v methanol:100 mM aqueous ammonium acetate	Half of the original extract volume
Tissue, whole organism or mammalian cell non-polar extract	Positive or negative ion	2:1 v/v 7.5 mM methanolic ammonium acetate:chloroform	Double the original extract volume
Biofluid polar extract	Positive ion	4:1 v/v methanol:H ₂ O. 0.25% (v/v) formic acid added.	5-10× the original extract volume
	Negative ion	4:1 v/v methanol:100 mM aqueous ammonium acetate	5-10× the original extract volume
Biofluid non-polar extract	Positive or negative ion	2:1 v/v 7.5 mM methanolic ammonium acetate:chloroform	5-10× the original extract volume

Table 3: Troubleshooting associated with nanoelectrospray ionisation and DIMS detection of samples.

Problem	Potential cause	Solution	Step where the problem originates	Step where the problem becomes evident
The nESI spray fails to initiate, is unstable or stops	Particulates in the sample	Centrifuge the resuspended sample at 4 °C prior to analysis. Take care not to disrupt the pellet during sample loading. Do not set the Triversa well plate temperature lower than 10 °C, which may encourage a precipitate to form.	4-5	10
	Sample too concentrated or too salty	Create (and analyse) a serial dilution of the resuspended sample to establish the most concentrated sample that gives a stable spray.	2	10
	Low sample viscosity causing sample to drip from the tip prior to analysis	In ChipSoft <i>Method Manager</i> , increase the ' <i>Volume of air to aspirate after sample</i> ' in 0.5 µL steps.	8	10
	Triversa nESI chip too close to MS source	In the <i>Spray Optimisation</i> window within ChipSoft <i>Method Manager</i> move the Triversa away from the MS source.	8	10
	Low nitrogen gas pressure in the supply to the Triversa	Check gas cylinder pressure	8	10
	nESI chip not loaded correctly	Remove and replace the chip	8	10
	Non-optimal Triversa pressure and voltage settings	Optimise the Triversa voltage and pressure settings as described in Box 1	8	10
Abnormally high nESI	Sample too concentrated or	Create (and analyse) a serial dilution of the resuspended sample to	2	10

spray current (>400nA)	too salty	establish the most concentrated sample that gives a stable spray.		
Abnormally low nESI spray current (<50nA)	Sample too dilute or insufficient modifier (formic acid or ammonium acetate) added	Create (and analyse) a serial dilution of the resuspended sample to establish the most concentrated sample that gives a stable spray. Check modifier addition (see Table 2).	2	10
nESI stable but poor stability of ions entering the MS	Triversa nESI chip too close to mass spectrometer source	In the <i>Spray Optimisation</i> window within ChipSoft <i>Method Manager</i> move the Triversa away from the MS source.	8	10
Good nESI current (100-300nA) but low ion intensity in MS	Triversa nESI chip too far away from mass spectrometer source	In the <i>Spray Optimisation</i> window within ChipSoft <i>Method Manager</i> move the Triversa closer to the MS source.	8	10
Triversa fails to pick up tips	Glue deposited on the mandrel from self-adhesive foil	Clean mandrel with xylene. Only use heat sealed foil to cover well plates.	6	10
Triversa running but MS failing to initiate	Contact closure failure	Check the contact closure cable connection between Triversa and mass spectrometer	8	10

Table 4: Average of the relative standard deviation (RSD) values of peak intensity measurements, from n repeated acquisitions of the same QC sample, using the spectral-stitching nESI DIMS method.

Instrument	RSD values of intensity measurements	Sample type	Number of peaks assessed	Number of repeated QC acquisitions (n)
LTQ FT™	8.1% (mean)	Polar liver extract ²³	17 endogenous compounds	3
LTQ FT Ultra™	8.2% (median) ^a	Polar liver extract ²⁹	All peaks in the dataset	80 ^b
Orbitrap Elite™	13.5% (median)	Polar serum extract	All peaks in the dataset	10
Q Exactive™	16.0% (median)	Polar human bronchial epithelial cell line extract	All peaks in the dataset	9

^aIncluding batch correction²⁹.

^bData acquired as 8 batches of 10 samples over a 7 day period.

Figure Legends

Figure 1 | Flow chart showing the full nESI DIMS spectral-stitching workflow. Definitions: Blank = extraction preparation done in the absence of sample; QC = quality control sample used to measure analytical reproducibility; PQN = probabilistic quotient normalisation; DIMS = direct infusion mass spectrometry.

Figure 2 | Schematic of the spectral-stitching nESI DIMS method for application on the Q Exactive™ Hybrid Quadrupole Orbitrap mass spectrometer. Data is collected as a series of 50 *m/z* wide SIM windows that overlap by 10 *m/z*. Windows are ‘stitched’ together to create a full mass spectrum. The optimised parameters for application of this method on the LTQ FT™, LTQ FT Ultra™, Q Exactive™ and Orbitrap Elite™ mass spectrometers are shown in Table 1.

Figure 3 | Recommended arrangement of samples in a 96-well plate for a spectral-stitching nESI DIMS experiment. Samples are each loaded as triplicates and are analysed in the order they appear in the plate (e.g. position A01, A02, A03, ..., H12). The plate begins with 6 ‘trial’ samples (A01-A06, typically 3x DIMS solvent (DS_T) and 3x QC samples (QC_T) that are used to check the nanoelectrospray stability of the DIMS system and the QC samples. This is followed by 3x extract blank (EB) samples. QC samples are injected after the EB, then after every 4 to 6 biological samples, and at the end of the plate. Biological samples (BS) across different sample classes (typically different biological phenotypes; shaded green and blue) are placed in a random order across the plate to minimise possible biases resulting from analysis order.

Figure 4 | Detection sensitivity of the spectral-stitching nESI DIMS method (blue) is superior to standard full scan analysis (dark orange). Data show the polar metabolites in a serum extract analysed on a Q ExactiveTM Hybrid Quadrupole Orbitrap mass spectrometer. Four QCs were each collected in triplicate and only peaks occurring in at least two-out-of-three technical replicates were retained. **A** shows the complete 50-620 m/z range; **C** is a zoomed in version of **A** covering the 355-360 m/z range; **B** and **C** show the average number of peaks detected over the 50-620 m/z and 355-360 m/z ranges, respectively. Error bars represent the standard deviation of four QC samples and circles represent the peak counts from each individual spectrum.

Figure 5 | Analysis of a human leukaemia cell line (HL60) lipid extract by the spectral-stitching nESI DIMS method in positive ion mode (**A**) and negative ion mode (**B**). Selected annotations and ppm m/z errors are shown. LPE – lysophosphatidylethanolamine; LPC – lysophosphatidylcholine; SM – sphingomyelin; Cer – ceramide; PE – phosphatidylethanolamine; PC – phosphatidylcholine; PS – phosphatidylserine; PI – phosphatidylinositol; TG – triacylglycerol; DG – diacylglycerol; Ac – acetate ion.

Figure 6 | PCA scores plot of polar extracts of *Daphnia pulex-pulicaria* collected by the spectral-stitching nESI DIMS method, in negative ion mode, on the LTQ FT UltraTM 74. F0 generation animals (light orange triangles) are separated from F1 generation animals (light blue squares) along PC2. Within the F0 and F1 generations there are clusters along the PC1 and PC2 axes. These are caused by biological differences of unknown origin. Most importantly the twenty eight QC samples (green circles) are tightly clustered indicating the high analytical reproducibility of the spectral-stitching method.

Figure 7 | Intensity measurements by spectral-stitch nESI DIMS method are consistent with NMR spectroscopic measurements. Comparison of LTQ FTTM spectral-stitching nESI DIMS data and NMR spectroscopic data measured on the same polar liver extract¹¹. Probabilistic quotient normalised (PQN⁶⁰) intensity measurements showing the NMR peak areas on the x-axis against DIMS peak intensity on the y-axis (the normalisation process makes the intensity values small [10^{-4}]). **A** tyrosine versus [tyrosine-H]⁺, **B** tyrosine versus [tyrosine+Cl]⁺, **C** creatine versus [creatine-H]⁺ and **D** phosphocreatine versus [phosphocreatine-H]⁺. Spearman's rank correlation coefficients (rho) and associated p-values are shown.

Figures

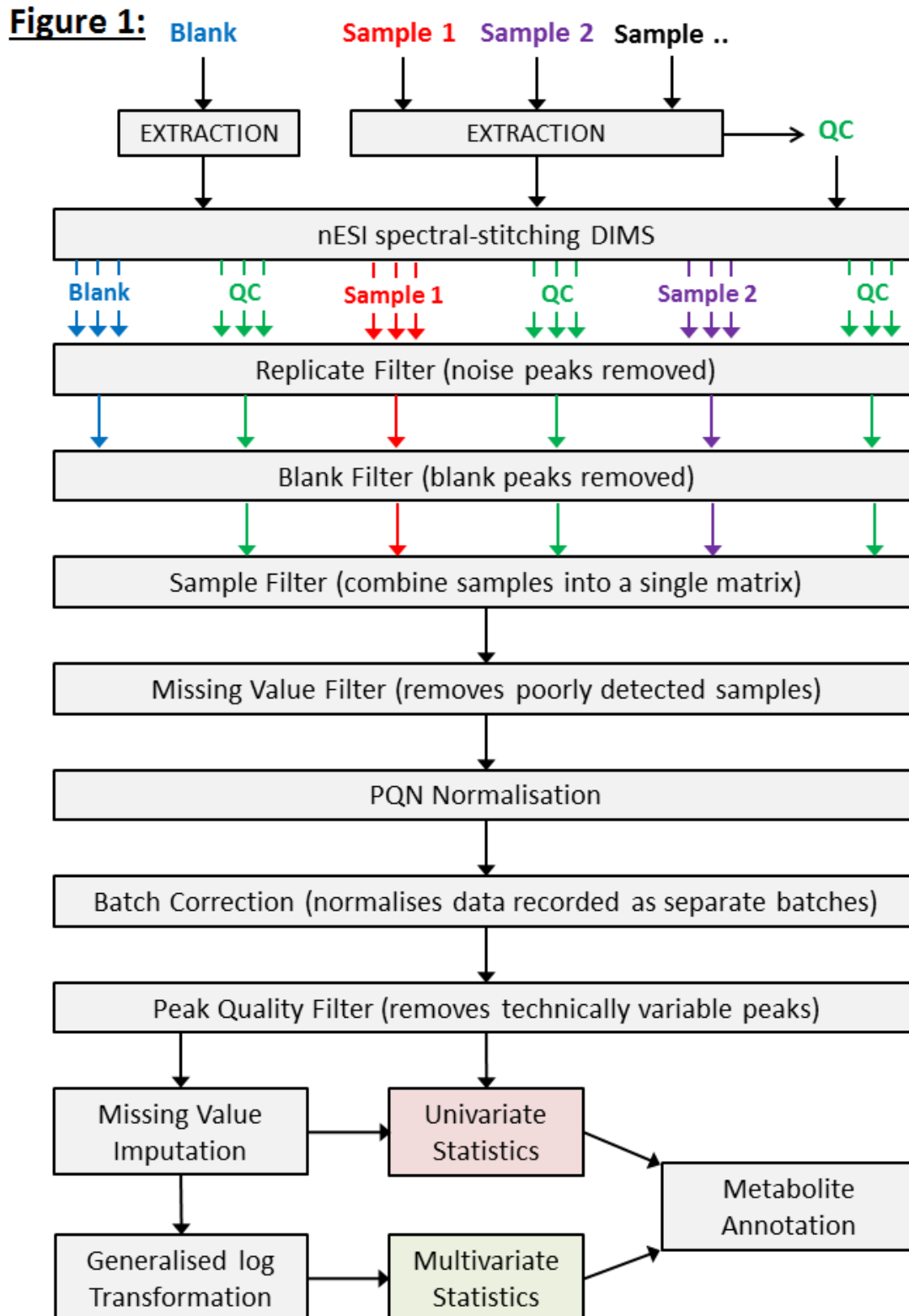


Figure 2:

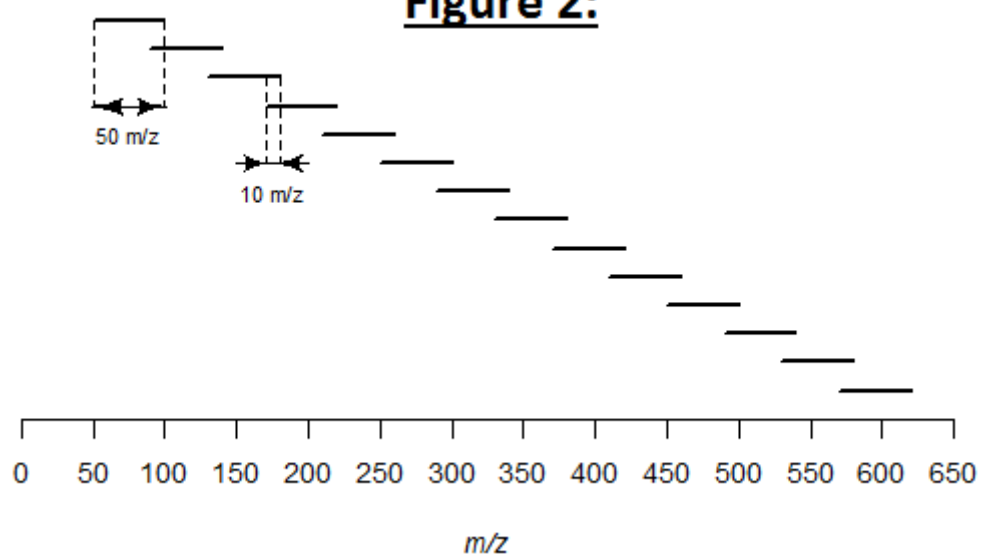


Figure 3

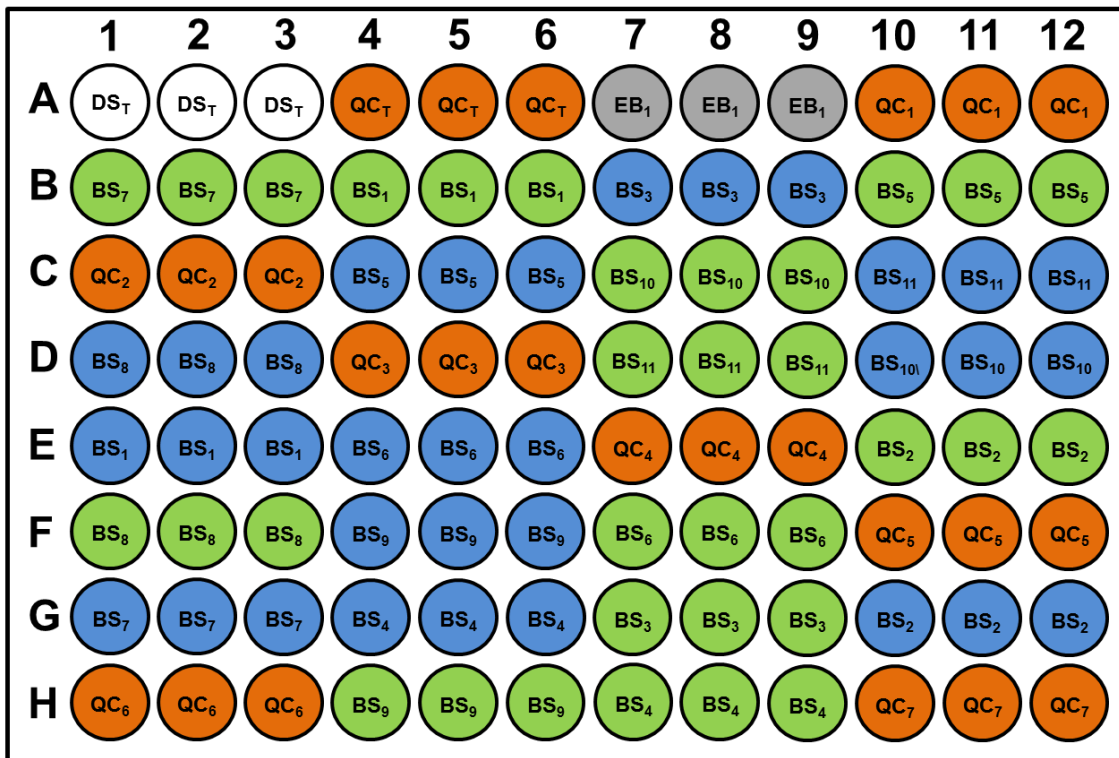


Figure 4

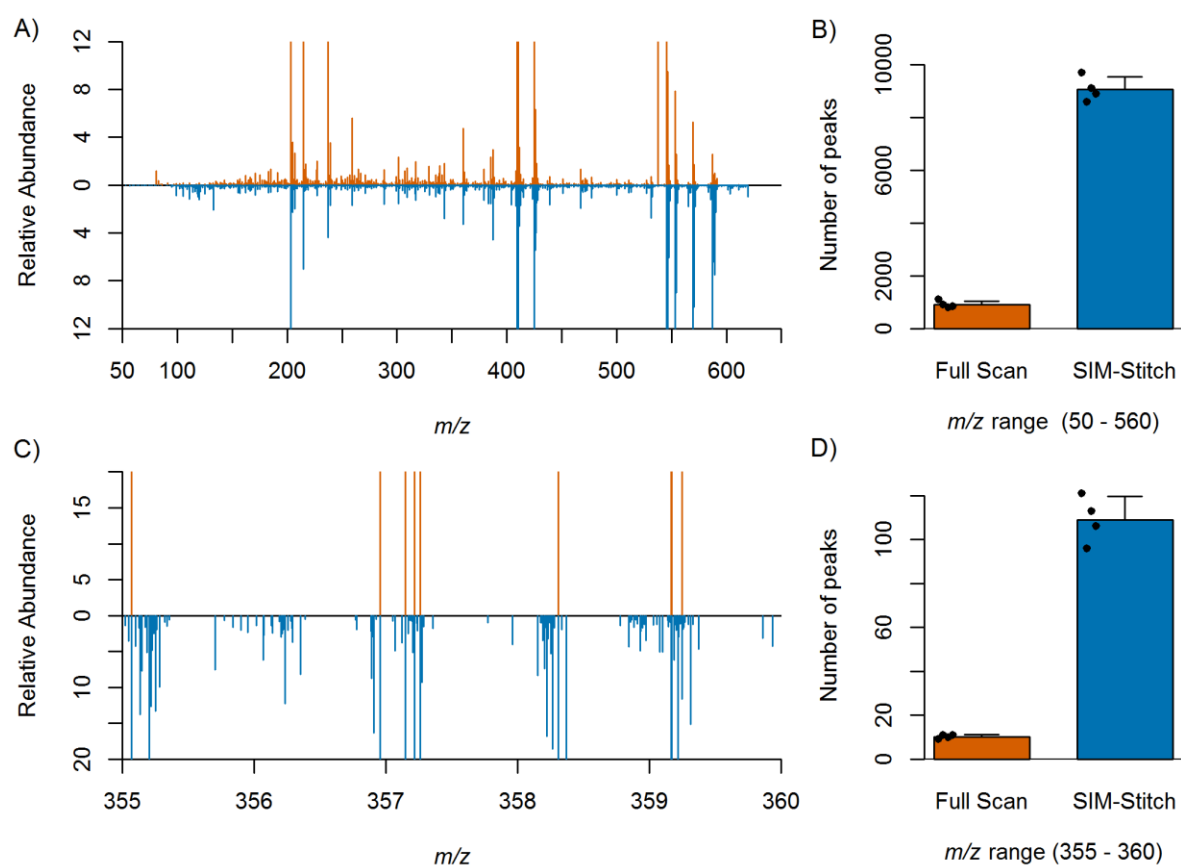


Figure 5

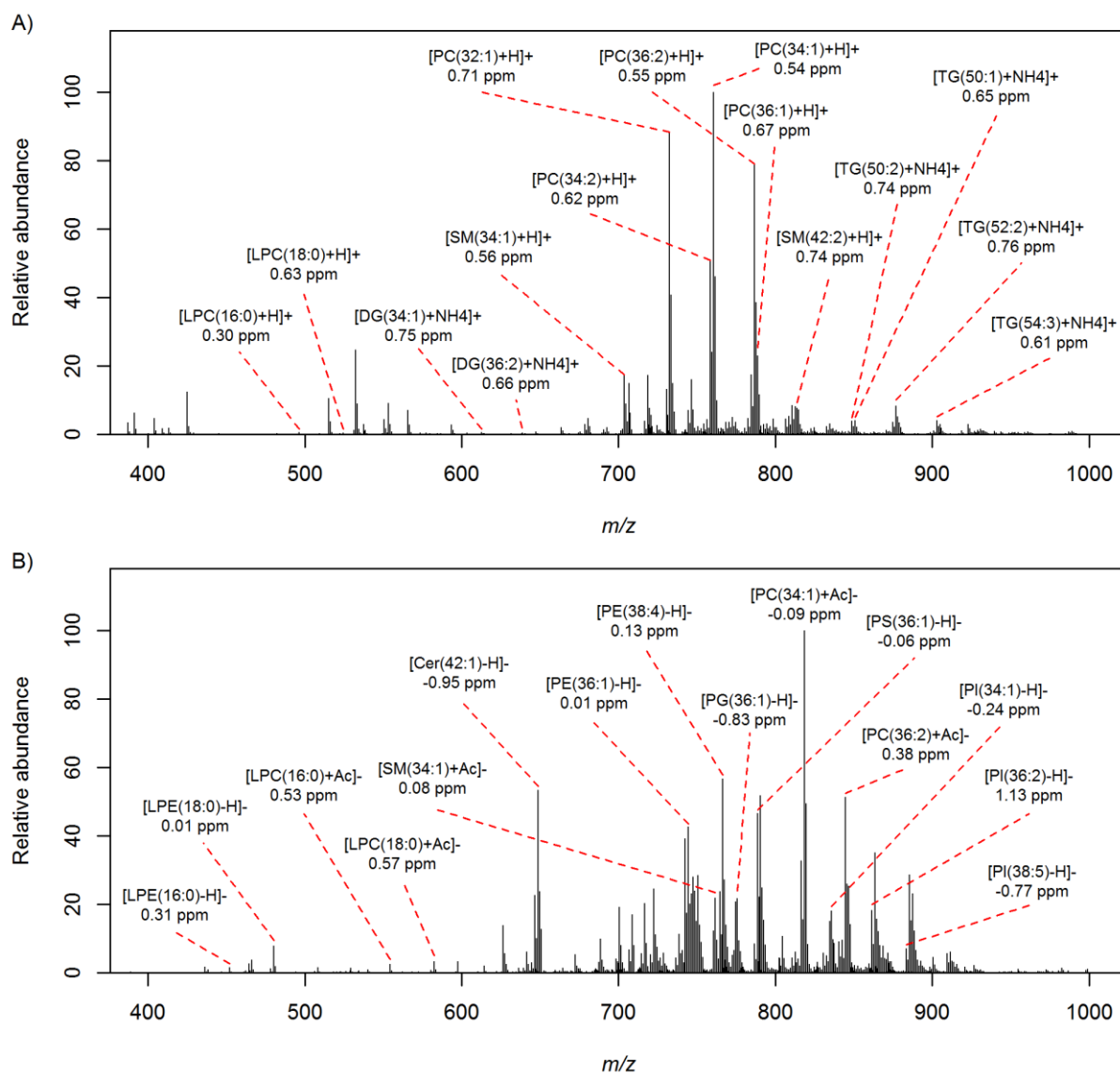


Figure 6

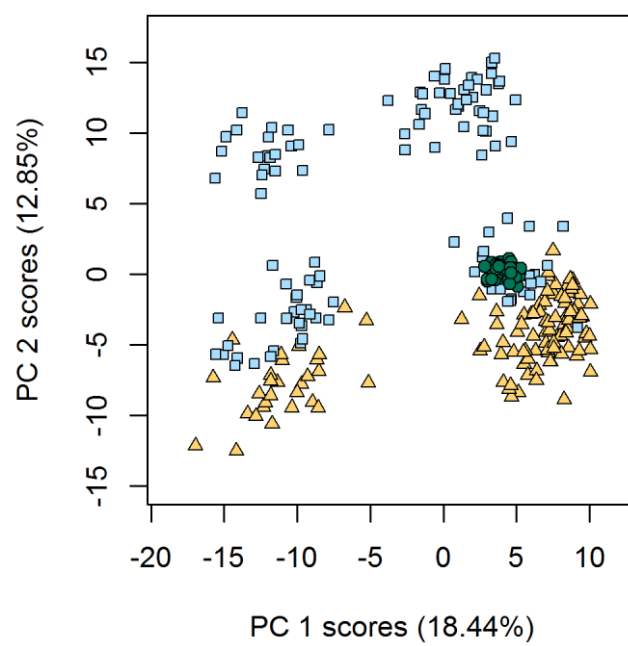
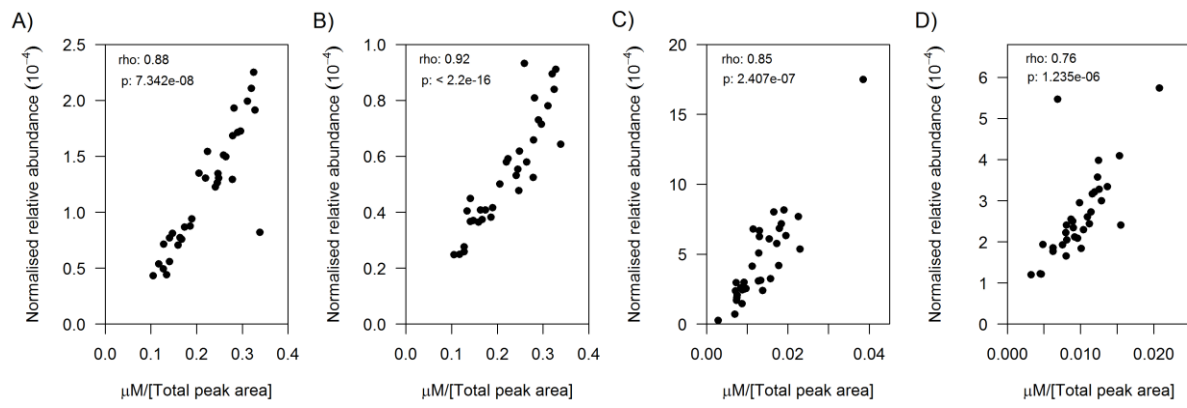


Figure 7



Supplementary Information

Supplementary Methods: Q ExactiveTM and Orbitrap EliteTM instrument method files (SI-1__Q_Exactive_Instrument_Method_Files.zip; SI-2__Orbitrap_Elite_Instrument_Method_Files.zip)

References

- 1 Fiehn, O. Metabolomics - the link between genotypes and phenotypes. *Plant Mol.Biol.* **48**, 155-171, doi:10.1023/a:1013713905833 (2002).
- 2 Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153-161, doi:10.1038/nrd728 (2002).
- 3 Han, X. L. & Gross, R. W. Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrometry Reviews* **24**, 367-412, doi:10.1002/mas.20023 (2005).
- 4 Wenk, M. R. The emerging field of lipidomics. *Nat. Rev. Drug Discov.* **4**, 594-610, doi:10.1038/nrd1776 (2005).
- 5 Quehenberger, O. *et al.* Lipidomics reveals a remarkable diversity of lipids in human plasma. *Journal of Lipid Research* **51**, 3299-3305, doi:10.1194/jlr.M009449 (2010).
- 6 Sreekumar, A. *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* **457**, 910-914, doi:10.1038/nature07762 (2009).
- 7 Southam, A. D. *et al.* Drug Redeployment to Kill Leukemia and Lymphoma Cells by Disrupting SCD1-Mediated Synthesis of Monounsaturated Fatty Acids. *Cancer Res*, doi:10.1158/0008-5472.can-15-0202 (2015).
- 8 Allen, J. *et al.* High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat. Biotechnol.* **21**, 692-696, doi:10.1038/nbt823 (2003).
- 9 Kamphorst, J. J. *et al.* Human Pancreatic Cancer Tumors Are Nutrient Poor and Tumor Cells Actively Scavenge Extracellular Protein. *Cancer Research* **75**, 544-553, doi:10.1158/0008-5472.can-14-2211 (2015).
- 10 Taylor, N. S. *et al.* A new approach to toxicity testing in *Daphnia magna*: application of high throughput FT-ICR mass spectrometry metabolomics. *Metabolomics* **5**, 44-58, doi:10.1007/s11306-008-0133-3 (2009).
- 11 Southam, A. D. *et al.* Metabolomics Reveals Target and Off-Target Toxicities of a Model Organophosphate Pesticide to Roach (*Rutilus rutilus*): Implications for Biomonitoring. *Environmental Science & Technology* **45**, 3759-3767, doi:10.1021/es103814d (2011).
- 12 Gruening, N.-M. *et al.* Pyruvate Kinase Triggers a Metabolic Feedback Loop that Controls Redox Metabolism in Respiring Cells. *Cell Metabolism* **14**, 415-427, doi:10.1016/j.cmet.2011.06.017 (2011).
- 13 Li, X., Gianoulis, T. A., Yip, K. Y., Gerstein, M. & Snyder, M. Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **143**, 639-650, doi:10.1016/j.cell.2010.09.048 (2010).
- 14 Kaelin, W. G., Jr. & McKnight, S. L. Influence of Metabolism on Epigenetics and Disease. *Cell* **153**, 56-69, doi:10.1016/j.cell.2013.03.004 (2013).
- 15 Dunn, W. B., Bailey, N. J. C. & Johnson, H. E. Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606-625, doi:10.1039/b418288j (2005).
- 16 Want, E. J., Cravatt, B. F. & Siuzdak, G. The expanding role of mass spectrometry in metabolite profiling and characterization. *Chembiochem* **6**, 1941-1951, doi:10.1002/cbic.200500151 (2005).
- 17 Jonsson, P. *et al.* High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Analytical Chemistry* **77**, 5635-5642, doi:10.1021/ac050601e (2005).
- 18 Theodoridis, G., Gika, H. G. & Wilson, I. D. LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *Trac-Trends Anal. Chem.* **27**, 251-260, doi:10.1016/j.trac.2008.01.008 (2008).
- 19 Ramautar, R., Somsen, G. W. & de Jong, G. J. CE-MS for metabolomics: Developments and applications in the period 2010-2012. *Electrophoresis* **34**, 86-98, doi:10.1002/elps.201200390 (2013).

- 20 Koulman, A. *et al.* High-resolution extracted ion chromatography, a new tool for metabolomics and lipidomics using a second-generation orbitrap mass spectrometer. *Rapid Commun. Mass Spectrom.* **23**, 1411-1418, doi:10.1002/rcm.4015 (2009).
- 21 Castrillo, J. I., Hayes, A., Mohammed, S., Gaskell, S. J. & Oliver, S. G. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**, 929-937, doi:10.1016/s0031-9422(02)00713-6 (2003).
- 22 Beckmann, M., Parker, D., Enot, D. P., Duval, E. & Draper, J. High-throughput, nontargeted metabolite fingerprinting using nominal mass flow injection electrospray mass spectrometry. *Nature Protocols* **3**, 486-504, doi:10.1038/nprot.2007.500 (2008).
- 23 Southam, A. D., Payne, T. G., Cooper, H. J., Arvanitis, T. N. & Viant, M. R. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Analytical Chemistry* **79**, 4595-4602, doi:10.1021/ac062446p (2007).
- 24 Weber, R. J. M., Southam, A. D., Sommer, U. & Viant, M. R. Characterization of Isotopic Abundance Measurements in High Resolution FT-ICR and Orbitrap Mass Spectra for Improved Confidence of Metabolite Identification. *Analytical Chemistry* **83**, 3737-3743, doi:10.1021/ac2001803 (2011).
- 25 Hop, C., Chen, Y. & Yu, L. J. Uniformity of ionization response of structurally diverse analytes using a chip-based nanoelectrospray ionization source. *Rapid Commun. Mass Spectrom.* **19**, 3139-3142, doi:10.1002/rcm.2182 (2005).
- 26 Kind, T. & Fiehn, O. Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **7**, doi:10.1186/1471-2105-7-234 (2006).
- 27 Sumner, L. W. *et al.* Proposed minimum reporting standards for chemical analysis. *Metabolomics* **3**, 211-221, doi:10.1007/s11306-007-0082-2 (2007).
- 28 Davidson, R. L., Weber, R. J. M., Liu, H., Sharma-Oates, A. & Viant, M. R. Galaxy-M: A galaxy workflow for processing and analysing direct infusion and liquid chromatography mass spectrometry-based metabolomics data. *Gigascience In Press* (2015).
- 29 Kirwan, J. A., Weber, R. J. M., Broadhurst, D. I. & Viant, M. R. Direct infusion mass spectrometry metabolomics dataset: a benchmark for data processing and quality control. *Scientific Data* **1**, doi:10.1038/sdata.2014.12 (2014).
- 30 Hao, J. *et al.* Statistical Correlations between NMR Spectroscopy and Direct Infusion FT-ICR Mass Spectrometry Aid Annotation of Unknowns in Metabolomics. *Analytical Chemistry* **88**, 2583-2589, doi:10.1021/acs.analchem.5b02889 (2016).
- 31 Lin, L. *et al.* Direct infusion mass spectrometry or liquid chromatography mass spectrometry for human metabolomics? A serum metabolomic study of kidney cancer. *Analyst* **135**, 2970-2978, doi:10.1039/c0an00265h (2010).
- 32 Annesley, T. M. Ion suppression in mass spectrometry. *Clinical Chemistry* **49**, 1041-1044, doi:10.1373/49.7.1041 (2003).
- 33 Schmidt, A., Karas, M. & Dulcks, T. Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? *J. Am. Soc. Mass Spectrom.* **14**, 492-500, doi:10.1016/s1044-0305(03)00128-4 (2003).
- 34 Yang, K. & Han, X. Accurate Quantification of Lipid Species by Electrospray Ionization Mass Spectrometry — Meets a Key Challenge in Lipidomics. *Metabolites* **1**, 21-40, doi:10.3390/metabo1010021 (2011).
- 35 Poynton, H. C. *et al.* Metabolomics of Microliter Hemolymph Samples Enables an Improved Understanding of the Combined Metabolic and Transcriptional Responses of *Daphnia magna* to Cadmium. *Environmental Science & Technology* **45**, 3710-3717, doi:10.1021/es1037222 (2011).
- 36 Gonzalez-Dominguez, R., Castilla-Quintero, R., Garcia-Barrera, T. & Luis Gomez-Ariza, J. Development of a metabolomic approach based on urine samples and direct infusion mass spectrometry. *Analytical Biochemistry* **465**, 20-27, doi:10.1016/j.ab.2014.07.016 (2014).

- 37 Lokhov, P. G. *et al.* Prediction of classical clinical chemistry parameters using a direct infusion mass spectrometry. *International Journal of Mass Spectrometry*, doi:<http://dx.doi.org/10.1016/j.ijms.2015.08.006> (2015).
- 38 Drury, N. E. *et al.* The effect of perhexiline on myocardial protection during coronary artery surgery: a two-centre, randomized, double-blind, placebo-controlled trial. *Eur. J. Cardio-Thorac. Surg.* **47**, 464-472, doi:10.1093/ejcts/ezu238 (2015).
- 39 Mirbahai, L. *et al.* Disruption of DNA Methylation via S-Adenosylhomocysteine Is a Key Process in High Incidence Liver Carcinogenesis in Fish. *Journal of Proteome Research* **12**, 2895-2904, doi:10.1021/pr400195u (2013).
- 40 Weber, R. J. M., Selander, E., Sommer, U. & Viant, M. R. A Stable-Isotope Mass Spectrometry-Based Metabolic Footprinting Approach to Analyze Exudates from Phytoplankton. *Mar. Drugs* **11**, 4158-4175, doi:10.3390/md11114158 (2013).
- 41 Southam, A. D. *et al.* Distinguishing between the metabolome and xenobiotic exposome in environmental field samples analysed by direct-infusion mass spectrometry based metabolomics and lipidomics. *Metabolomics* **10**, 1050-1058, doi:10.1007/s11306-014-0693-3 (2014).
- 42 Sturla, S. J. *et al.* Systems Toxicology: From Basic Research to Risk Assessment. *Chemical Research in Toxicology* **27**, 314-329, doi:10.1021/tx400410s (2014).
- 43 Van Aggelen, G. *et al.* Integrating Omic Technologies into Aquatic Ecological Risk Assessment and Environmental Monitoring: Hurdles, Achievements, and Future Outlook. *Environmental Health Perspectives* **118**, 1-5, doi:10.1289/ehp.0900985 (2010).
- 44 Parsons, H. M., Ekman, D. R., Collette, T. W. & Viant, M. R. Spectral relative standard deviation: a practical benchmark in metabolomics. *Analyst* **134**, 478-485, doi:10.1039/b808986h (2009).
- 45 Sellick, C. A., Hansen, R., Stephens, G. M., Goodacre, R. & Dickson, A. J. Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nature Protocols* **6**, 1241-1249, doi:10.1038/nprot.2011.366 (2011).
- 46 Taylor, N. S., Weber, R. J. M., White, T. A. & Viant, M. R. Discriminating between Different Acute Chemical Toxicities via Changes in the Daphnid Metabolome. *Toxicol. Sci.* **118**, 307-317, doi:10.1093/toxsci/kfq247 (2010).
- 47 Want, E. J. *et al.* Global metabolic profiling of animal and human tissues via UPLC-MS. *Nature Protocols* **8**, 17-32, doi:10.1038/nprot.2012.135 (2013).
- 48 Want, E. J. *et al.* Global metabolic profiling procedures for urine using UPLC-MS. *Nature Protocols* **5**, 1005-1018, doi:10.1038/nprot.2010.50 (2010).
- 49 Dunn, W. B. *et al.* Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nature Protocols* **6**, 1060-1083, doi:10.1038/nprot.2011.335 (2011).
- 50 Wu, H., Southam, A. D., Hines, A. & Viant, M. R. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Analytical Biochemistry* **372**, 204-212, doi:10.1016/j.ab.2007.10.002 (2008).
- 51 Lin, C. Y., Wu, H. F., Tjeerdema, R. S. & Viant, M. R. Evaluation of metabolite extraction strategies from tissue samples using NMR metabolomics. *Metabolomics* **3**, 55-67, doi:10.1007/s11306-006-0043-1 (2007).
- 52 Han, X., Yang, K. & Gross, R. W. Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses. *Mass Spectrometry Reviews* **31**, 134-178, doi:10.1002/mas.20342 (2012).
- 53 Bligh, E. G. & Dyer, W. J. A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION. *Canadian Journal of Biochemistry and Physiology* **37**, 911-917 (1959).
- 54 Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of Lipid Research* **49**, 1137-1146, doi:10.1194/jlr.D700041-JLR200 (2008).

- 55 Schultz, G. A., Corso, T. N., Prosser, S. J. & Zhang, S. A fully integrated monolithic microchip electrospray device for mass spectrometry. *Analytical Chemistry* **72**, 4058-4063, doi:10.1021/ac000325y (2000).
- 56 Zhang, L. K., Rempel, D., Pramanik, B. N. & Gross, M. L. Accurate mass measurements by Fourier transform mass spectrometry. *Mass Spectrometry Reviews* **24**, 286-309, doi:10.1002/mas.20013 (2005).
- 57 Payne, T. G., Southam, A. D., Arvanitis, T. N. & Viant, M. R. A Signal Filtering Method for Improved Quantification and Noise Discrimination in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Data. *J. Am. Soc. Mass Spectrom.* **20**, 1087-1095, doi:10.1016/j.jasms.2009.02.001 (2009).
- 58 Kozhinov, A. N., Zhurov, K. O. & Tsybin, Y. O. Iterative Method for Mass Spectra Recalibration via Empirical Estimation of the Mass Calibration Function for Fourier Transform Mass Spectrometry-Based Petroleomics. *Analytical Chemistry* **85**, 6437-6445, doi:10.1021/ac400972y (2013).
- 59 Lange, O., Damoc, E., Wieghaus, A. & Makarov, A. Enhanced Fourier transform for Orbitrap mass spectrometry. *International Journal of Mass Spectrometry* **369**, 16-22, doi:10.1016/j.ijms.2014.05.019 (2014).
- 60 Dieterle, F., Ross, A., Schlotterbeck, G. & Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in H-1 NMR metabonomics. *Analytical Chemistry* **78**, 4281-4290, doi:10.1021/ac051632c (2006).
- 61 Kirwan, J. A., Broadhurst, D. I., Davidson, R. L. & Viant, M. R. Characterising and correcting batch variation in an automated direct infusion mass spectrometry (DIMS) metabolomics workflow. *Analytical and Bioanalytical Chemistry* **405**, 5147-5157, doi:10.1007/s00216-013-6856-7 (2013).
- 62 Hrydziusko, O. & Viant, M. R. Missing values in mass spectrometry based metabolomics: an undervalued step in the data processing pipeline. *Metabolomics* **8**, S161-S174, doi:10.1007/s11306-011-0366-4 (2012).
- 63 Parsons, H. M., Ludwig, C., Gunther, U. L. & Viant, M. R. Improved classification accuracy in 1-and 2-dimensional NMR metabolomics data using the variance stabilising generalised logarithm transformation. *BMC Bioinformatics* **8**, 16, doi:10.1186/1471-2105-8-234 (2007).
- 64 van den Berg, R. A., Hoefsloot, H. C. J., Westerhuis, J. A., Smilde, A. K. & van der Werf, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *Bmc Genomics* **7**, doi:10.1186/1471-2164-7-142 (2006).
- 65 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
- 66 Good, P. L. *Permutation, Parametric, and Bootstrap Tests of Hypotheses*. (Springer, 2005).
- 67 Szymanska, E., Saccenti, E., Smilde, A. K. & Westerhuis, J. A. Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies. *Metabolomics* **8**, 3-16, doi:10.1007/s11306-011-0330-3 (2012).
- 68 Dunn, W. B. *et al.* Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. *Metabolomics* **9**, S44-S66, doi:10.1007/s11306-012-0434-4 (2013).
- 69 Weber, R. J. M. & Viant, M. R. MI-Pack: Increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways. *Chemometrics Intell. Lab. Syst.* **104**, 75-82, doi:10.1016/j.chemolab.2010.04.010 (2010).
- 70 Kind, T. & Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* **8**, doi:10.1186/1471-2105-8-105 (2007).
- 71 Haug, K. *et al.* MetaboLights-an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Research* **41**, D781-D786, doi:10.1093/nar/gks1004 (2013).
- 72 Sud, M. *et al.* Metabolomics Workbench: An international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools. *Nucleic Acids Res*, doi:10.1093/nar/gkv1042 (2015).

- 73 Bennett, B. D., Yuan, J., Kimball, E. H. & Rabinowitz, J. D. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. *Nature Protocols* **3**, 1299-1311, doi:10.1038/nprot.2008.107 (2008).
- 74 Taylor, N. S. *et al.* Metabolomics confirms that dissolved organic carbon mitigates copper toxicity. *Environmental toxicology and chemistry / SETAC*, doi:10.1002/etc.3206 (2015).